NAVAL
POSTGRADUATE
SCHOOL

MONTEREY, CALIFORNIA

DISSERTATION

GALERKIN OPTIMAL CONTROL

by

Randy Boucher

December 2014

Dissertation Supervisor: Wei Kang

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.

1. AGENCY USE ONLY (Leave blank)	**2. REPORT DATE** December 2014	**3. REPORT TYPE AND DATES COVERED** Dissertation – October 13 - December 14

4. TITLE AND SUBTITLE: GALERKIN OPTIMAL CONTROL	**5. FUNDING NUMBERS**
6. AUTHOR(S): Randy Boucher	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000	**8. PERFORMING ORGANIZATION REPORT NUMBER**
9.SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**

11. SUPPLEMENTARY NOTES: The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited	**12b. DISTRIBUTION CODE** A

13. ABSTRACT (maximum 200 words)

A Galerkin-based family of numerical formulations is presented for solving nonlinear optimal control problems. This dissertation introduces a family of direct methods that calculate optimal trajectories by discretizing the system dynamics using Galerkin numerical techniques and approximate the cost function with Gaussian quadrature. In this numerical approach, the analysis is based on L^2-norms. An important result in the theoretical foundation is that the feasibility and consistency theorems are proved for problems with continuous and/or piecewise continuous controls. Galerkin methods may be formulated in a number of ways that allow for efficiency and/or improved accuracy while solving a wide range of optimal control problems with a variety of state and control constraints. Numerical formulations using Lagrangian and Legendre test functions are derived. One formulation allows for a weak enforcement of boundary conditions, which imposes end conditions only up to the accuracy of the numerical approximation itself. Additionally, the multi-scale formulation can reduce the dimension of multi-scale optimal control problems, those in which the states and controls evolve on different timescales. Finally, numerical examples are shown to demonstrate the versatile nature of Galerkin optimal control.

14. SUBJECT TERMS Galerkin, Pseudospectral, Optimal Control, Constrainted Optimization	**15. NUMBER OF PAGES** 217
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UU

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

GALERKIN OPTIMAL CONTROL

Randy Boucher
Lieutenant Colonel, United States Army
B.S., Boston University, 1997
M.S., University of Washington, 2006

Submitted in partial fulfillment of the
requirements for the degree of

**DOCTOR OF PHILOSOPHY IN
APPLIED MATHEMATICS**

from the

**NAVAL POSTGRADUATE SCHOOL
December 2014**

Author: Randy Boucher

Approved By: Wei Kang
Professor
Department of Appl. Math.
Dissertation Supervisor

Chris L. Frenzen
Professor
Department of Appl. Math.

Francis X. Giraldo
Professor
Department of Appl. Math.

Arthur J. Krener
Professor
Department of Appl. Math.

I. Michael Ross
Professor
Department of Mech. & Aero.

Approved By: Craig Rasmussen, Professor and Chair, Department of Appl. Math.

Approved By: Doug Moses, Vice Provost for Academic Affairs

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

A Galerkin-based family of numerical formulations is presented for solving nonlinear optimal control problems. This dissertation introduces a family of direct methods that calculate optimal trajectories by discretizing the system dynamics using Galerkin numerical techniques and approximate the cost function with Gaussian quadrature. In this numerical approach, the analysis is based on L^2-norms. An important result in the theoretical foundation is that the feasibility and consistency theorems are proved for problems with continuous and/or piecewise continuous controls. Galerkin methods may be formulated in a number of ways that allow for efficiency and/or improved accuracy while solving a wide range of optimal control problems with a variety of state and control constraints. Numerical formulations using Lagrangian and Legendre test functions are derived. One formulation allows for a weak enforcement of boundary conditions, which imposes end conditions only up to the accuracy of the numerical approximation itself. Additionally, the multi-scale formulation can reduce the dimension of multi-scale optimal control problems, those in which the states and controls evolve on different timescales. Finally, numerical examples are shown to demonstrate the versatile nature of Galerkin optimal control.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to begin by thanking my wife, Jill. You are my compass and my strength. Thank you for keeping me anchored to all things that truly matter in life. Thanks also to my beautiful children, Leah, Dominic and Stella, who greet me with love and a smile at the end of each day.

I would also like to acknowledge my wonderful team of researchers. Thanks especially to Dr. Wei Kang for your patience and direction. You always made me feel like a colleague although you are one of many giants on whose shoulders I stand. Thanks also to the members of my committee: Dr. Frank Giraldo, Dr. Chris Frenzen, Dr. Mike Ross and Dr. Arthur Krener. Each of you planted great seeds of knowledge within me for which I am forever grateful.

I have encountered many great people along this journey. Thank you to all of my friends at NPS for lending an ear and raising a pint!

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 1:
INTRODUCTION

The last two decades have proven to be a time of active research for numerical methods for optimal control. Particularly, direct collocation methods, such as pseudospectral (PS) methods, have received much attention [1–8]. PS methods produce accurate solutions on a wide variety of optimal control problems. Two recent highlights are the successful use of the Legendre PS method for the first ever zero-propellant attitude maneuver of the International Space Station [4] and the first ever minimum-time rotational maneuver performed in orbit by a NASA space telescope called TRACE [8]. In the Legendre PS method [1, 2, 5, 6, 9, 10], the problem is discretized at the Legendre-Gauss-Lobatto (LGL) points, Legendre-Gauss-Radau (LGR) points or Legendre-Gauss (LG) points. The states are approximated with globally interpolating Lagrange polynomials and the cost function is typically approximated using Gaussian quadrature rule. Other variants of the PS method include the Chebyshev PS method [11], the PS knotting method [12] and the Bellman method [13]. The Legendre PS method will be outlined in Chapter 3 of this dissertation, preceded by a review of mathematical topics in Chapter 2.

While PS methods have shown to be good all-round methods for solving nonlinear optimal control problems, approximating the derivative of a function using a standard PS differentiation matrix (such as the Legendre PS differentiation matrix) may introduce errors into the approximation. Chapter 3 highlights this issue with the use of Jackson's Theorem. Additionally, Chapter 3 will motivate the use of the weak integral formulation to approximate the system's dynamics. This leads to the creation of a family of Galerkin-based formulations called, "Galerkin optimal control."

The family of methods proposed in this dissertation are derived from Galerkin numerical techniques that have been developed for numerical solutions to differential equations since the early 1970s [14–16]. In addition to the family of Galerkin optimal control formulations that are presented, this dissertation highlights important theorems that prove method feasibility and consistency for problems with continuous and/or *piecewise continuous* controls.

The base Galerkin optimal control method is outlined in Chapter 4, where feasibility and convergence theorems are presented. Chapter 5 presents a review of additional Galerkin-based formulations and strategies such as the use element-based Galerkin techniques and a multi-scale approach. Lastly, modifications to the method such as over-integration and the use of various quadrature rules are offered to improve computational efficiency and/or increase accuracy of the solutions.

The remainder of the dissertation is organized as follows: Chapter 6 presents a Petrov-Galerkin optimal control approach to discretizing the system dynamics; in place of Lagrange polynomial test functions integrated into the base formulation, a set of Legendre polynomials are used. Improved feasibility and convergence theorems are presented. Chapter 7 demonstrates the versatile nature of the Galerkin optimal control formulations by considering a number of example problems. Lastly, Chapter 8, highlights the potential for Galerkin optimal control in solving a wide range of real-world optimal control problems with a variety of state and control constraints. Additionally, Chapter 8 discusses areas of future research.

CHAPTER 2: MATHEMATICAL BACKGROUND

2.1. Optimal Control

Optimal control has a rich history that dates back to 1696, when Johann Bernoulli posed the *bachristochrone problem* in the Acta Eruditorum to [17, 18] "*the most astute mathematicians of the world.*" The bachristochrone problem was the following:

> *If in the vertical plane two points A and B are given, then it is required to specify the orbit AMB of the movable point M, along which it, starting from A, and under the influence of its own weight, arrives at B in the shortest possible time.* [19]

In addition to Johann Bernoulli, other mathematical giants living in Europe at this time, such as Newton, Leibniz and Johann's brother, Jacob Bernoulli [19] (all considered "Men of Mathematics" by Bell [20]), solved the bachristochrone problem. Later, Euler invented a method for solving such problems (with mathematical underpinnings created by Lagrange), known today as the foundations of the calculus of variations. The standard calculus of variations problems is of the form [21]

$$\text{minimize } J[y(\cdot)] = \int_{t_0}^{t_f} F(t, y(t), \dot{y}(t))dt, \tag{2.1}$$

$$\text{subject to } y(t_0) = y_0 \text{ and } y(t_f) = y_f, \tag{2.2}$$

where J acts on a set of functions and is called a *functional*. Notice that problem (2.1)–(2.2) may be written in the equivalent optimal control problem form

$$\text{minimize } J[x(\cdot), u(\cdot)] = \int_{t_0}^{t_f} F(x(t), u(t))dt, \tag{2.3}$$

$$\text{subject to } x(t_0) = [t_0, y_0]^T,\ x(t_f) = [t_f, y_f]^T \text{ and } \dot{x}(t) = [1, u(t)]^T,\ t \in [t_0, t_f], \tag{2.4}$$

by renaming the variables t and y as $t = y_1$ and $y = y_2$, and creating the new vector $x = [y_1, y_2]^T$.

Over 250 years later, after many periods of active research in the field of calculus of variations, the Russian mathematician Lev Semenovich Pontryagin made a giant leap forward. In 1956, Pontryagin and his group established the optimal control theory [22, 23]. In contrast to standard calculus of variations problems of the form (2.1)–(2.2), or equivalent form (2.3)–(2.4), it was shown that optimal control theory was well suited to handle discontinuous solutions, $u(t)$. Additionally, Pontryagin established that problems of optimal control involved the minimization of a functional over a set of function pairs, $t \mapsto (x, u) \in \mathbb{R}^{N_x} \times \mathbb{R}^{N_u}$, subject to the dynamical constraint

$$\dot{x}(t) = f(x(t), u(t)),$$

where $f : \mathbb{R}^{N_x} \times \mathbb{R}^{N_u} \to \mathbb{R}^{N_x}$ and $u(t)$ is a control function. It was soon realized that this new theory of optimal control was well suited to solve many complex problems (that the calculus of variations could not). Over the last half century, optimal control theory has been developed into an extremely powerful tool that has touched many areas of mathematics, science and engineering. Consider the following general problem of optimal control.

2.1.1. The Optimal Control Problem

Determine the state-control function pair, $t \mapsto (x, u) \in \mathbb{R}^{N_x} \times \mathbb{R}^{N_u}$, that minimizes the cost functional

$$J[x(\cdot), u(\cdot)] = \int_{t_0}^{t_f} F(x(t), u(t))dt + E(x(t_f)), \tag{2.5}$$

subject to the dynamics,

$$\dot{x}(t) = f(x(t), u(t)), \tag{2.6}$$

initial conditions,

$$x(t_0) = x_0, \tag{2.7}$$

at specified time, t_0, and endpoint conditions,

$$e(x(t_f)) = 0, \tag{2.8}$$

where the running (or Lagrange) cost $F : \mathbb{R}^{N_x} \times \mathbb{R}^{N_u} \to \mathbb{R}$, the endpoint (or Mayer) cost, $E : \mathbb{R}^{N_x} \times \mathbb{R}^{N_x} \to \mathbb{R}$, $f : \mathbb{R}^{N_x} \times \mathbb{R}^{N_u} \to \mathbb{R}^{N_x}$ and $e : \mathbb{R}^{N_x} \times \mathbb{R}^{N_x} \to \mathbb{R}^{N_e}$, are Lipschitz continuous with respect to their arguments. A set of necessary conditions must be met in order to find candidate solutions to problem (2.5)–(2.8). Pontryagin's Minimum Principle provides the necessary framework.

2.1.2. Pontryagin's Minimum Principle

Pontryagin's Minimum Principle was proved by Pontryagin in 1956 [22, 23]. It provides conditions that must be met in order for a solution to be considered optimal. As with the calculus of variations both necessary and sufficient conditions for optimality may be established. Although sufficient conditions are beyond the scope of this dissertation (see [24]), first order necessary conditions will be outlined with help from the calculus of variations.

2.1.2.1. Calculus of Variations

In the calculus of variations, problems of the form (2.1)–(2.2) are solved by considering the variation of J, or ΔJ, given by

$$\Delta J[y^*, y] = J[y] - J[y^*]$$

where y^* is the minimizing curve, and y are all other admissible curves. For y^* to be a minimizing curve it is necessary that $\Delta J[y^*, y] \geq 0$. Additionally, if all the first order terms are collected in the expansion of ΔJ, it is necessary that this collection (called the first variation or shown symbolically as $\delta J[y^*]$) must be equal to zero [21]. This same approach may be used to define the optimality conditions for problem (2.5)–(2.8).

2.1.2.2. Necessary Conditions

In order to apply a variational approach to problem (2.5)–(2.8), consider the augmented functional,

$$J_a[x(\cdot), u(\cdot), \lambda(\cdot), \nu(\cdot)] = \int_{t_0}^{t_f} \big(F(x(t), u(t)) + \lambda^T(f(x(t), u(t)) - \dot{x}(t))\big)\, dt \\ + E(x(t_f)) - \nu^T e(x(t_f)),$$

where $\lambda(t) \in \mathbb{R}^{N_x}$ and $\nu(t) \in \mathbb{R}^{N_e}$ are Lagrange multipliers, and $\lambda(t)$ is typically given the name costate or adjoint covector. As in the calculus of variations approach, considering the first variation, $\delta J_a[u^*] = 0$, a set of necessary conditions can be obtained [21, 25–27]

$$\dot{x}(t) = \frac{\partial H}{\partial \lambda}, \tag{2.9}$$

$$\dot{\lambda}(t) = -\frac{\partial H}{\partial x}, \tag{2.10}$$

where the Hamiltonian, H, is given by

$$H(x(t), u(t), \lambda(t)) = F(x(t), u(t)) + \lambda^T f(x(t), u(t)). \tag{2.11}$$

Additionally, the Hamiltonian (2.11) reaches its minimum with respect to u at $u = u^*$. This is called the Hamiltonian Minimization Condition and can be expressed as

$$u^* = \arg\max_{u \in \mathbb{U}} H(x(t), u(t), \lambda(t)), \tag{2.12}$$

where $\mathbb{U}$ defines a region of feasible control. Finally, the following conditions must be satisfied on the boundary

$$\lambda(t_f) = \frac{\partial \bar{E}}{\partial x_f}, \tag{2.13}$$

$$H(t_f) = -\frac{\partial \bar{E}}{\partial t_f}, \tag{2.14}$$

$$e(x(t_f)) = 0, \tag{2.15}$$

where the endpoint Lagrangian, $\bar{E}$, is given by

$$\bar{E}(x(t_f), \nu) = E(x(t_f)) + \nu^T e(x(t_f)). \tag{2.16}$$

Equations (2.9), (2.10), (2.12) and (2.13)–(2.15) provide the first-order necessary conditions for optimality and create the framework for Pontryagin's Minimum Principle.

2.1.2.3. Pontryagin's Minimum Principle

Lemma 2.1 (Pontryagin's Minimum Principle)**.** [26] *Let, $(x^*(t), u^*(t))$, be a solution to problem (2.5)–(2.8). Then in order for $x^*(t)$ and $u^*(t)$ to be optimal, it is necessary that there exists a costate, λ, and covector, ν, that satisfies conditions (2.9), (2.10), (2.12) and (2.13)–(2.15).*

Remark 2.1. *For problem (2.5)–(2.8), with added path condition, the following mixed state-control inequality path constraint is included,*

$$h(x(t), u(t)) \leq 0, \tag{2.17}$$

where $h : \mathbb{R}^{N_x} \times \mathbb{R}^{N_u} \to \mathbb{R}^{N_h}$ is Lipschitz continuous with respect to x and u.

With the addition of the path constraint, candidate solutions to problem (2.5)–(2.8) and (2.17) can be found by solving the nonlinear programing (NLP) problem

$$u^* = \arg \max_{u \in \mathbb{U}(x)} \bar{H}(x(t), u(t), \lambda(t), \mu(t)),$$

where the constraint set, $\mathbb{U} \subseteq \mathbb{R}^{N_u}$*, is given by*

$$\mathbb{U}(x) = \{u \in \mathbb{U} | h(x(t), u(t)) \leq 0, x \in \mathbb{R}^{N_x}, t \in [t_0, t_f]\}.$$

The augmented Hamiltonian (or Lagrangian of the Hamiltonian), $\bar{H}$*, is given by*

$$\bar{H}(x(t), u(t), \lambda(t), \mu(t)) = H(x(t), u(t), \lambda(t)) + \mu^T h(x(t), u(t)),$$

where $\mu(t) \in \mathbb{R}^{N_h}$ *are Lagrange multipliers. The modified set of necessary conditions are* [27, 28]

$$\dot{x}(t) = \frac{\partial \bar{H}}{\partial \lambda}, \tag{2.18}$$

$$\dot{\lambda}(t) = -\frac{\partial \bar{H}}{\partial x}, \tag{2.19}$$

along with the following conditions on the boundary

$$\lambda(t_f) = \frac{\partial \bar{E}}{\partial x_f}, \tag{2.20}$$

$$\bar{H}(t_f) = -\frac{\partial \bar{E}}{\partial t_f}, \tag{2.21}$$

$$e(x(t_f)) = 0. \tag{2.22}$$

Additionally, the complementary (slackness) condition,

$$\mu_i \begin{cases} \leq 0, & h_i(x(t), u(t)) = 0, \\ = 0, & h_i(x(t), u(t)) < 0, \end{cases} \tag{2.23}$$

must be satisfied.

Equations (2.18)–(2.23) provide the first-order necessary conditions for optimality for problem (2.5)–(2.8) and (2.17).

Although Pontryagin provided a framework for finding candidate optimal solutions, many problems of optimal control are too difficult to solve analytically. It is easy to see the difficulty in solving the $2N_x$ Hamiltonian system of differential equations (2.9)–(2.10) or (2.18)–(2.19). For this reason, numerically methods have become extremely important in solving optimal control problems.

2.1.3. Numerical Methods for Optimal Control

Many numerical techniques have been investigated for solving optimal control problems since Pontryagin proved the Minimum Principle in 1956. These optimal control methods take two main forms, indirect and direct. Recent surveys of these techniques are provided by Betts [29, 30], Trélat [31] and Ross [32] and a historical perspective by Stryk et al. [33]. Indirect methods (such as the shooting and multiple shooting methods) solve Pontryagin's necessary conditions for optimality. Although these methods have been shown to solve a wide range of problems with great accuracy, they prove to be difficult to implement, due to the knowledge of the calculus of variations required and the difficulty of providing good initial guesses. In contrast, the direct methods (such as Euler, Runge-Kutta and collocation methods) discretize the cost function, problem dynamics, etc, at specified time points. Due to the fact that direct methods require no knowledge of the necessary conditions for optimality, and the accuracies that may be obtained, they have recently gained much attention. Of the direct methods, specifically the global orthogonal collocation methods (a.k.a. pseudospectral methods) have proven to solve difficult problems with great accuracy [4, 8, 34, 35] after becoming an actively researched topic in the 1990s by Elnagar et al. [1] and Fahroo et al. [2].

Pseudospectral (PS) methods for optimal control discretize the problem at specified nodes, called collocation points. Due to the properties of the orthogonal family of collocation points (such as those found via the Legendre or Chebyshev polynomial basis) approximations converge at spectral rates [6]. The most widely used Legendre PS method [1, 5, 6, 9, 10] is based on the LGL points [36]. However, the Legendre PS method may be based upon LGR or LG nodes as well [36, 37]. PS methods for optimal control have been formally implemented in the MATLAB-based software package DIDO [38] and NASA's Fortran-based package OTIS [39].

There are four parts to the numerical solution to an optimal control problem using a PS method: discretization of the system dynamics, discretization of the state-control constraints, integration of the cost function and solving the nonlinear program (NLP). The mathematical background associated with the first three steps will be discussed in the following sections. Spectral methods are attractive for discretizing the problem's dynamics due to their superior accuracy. Two global spectral methods, collocation and Galerkin, will be outlined in Section 2.3.1. Additionally, Galerkin methods may be formulated as element-based methods. These local spectral element methods will be outlined in Section 2.3.2. A fundamental task in the formulation of these global and local methods is the selection of good discretization points and the use of interpolating functions. Both will be discussed in detail in Section 2.2. Finally, numerical integration, or quadrature, is typically used to integrate the cost function and will also be outlined in Section 2.2.

The resulting NLP can be solved by using a commercial sequential quadratic programming (SQP) software packages such as dense NLP solver NPSOL [40] and sparse NLP solvers SNOPT [41, 42] and SPRNLP [43]. A feasible solution can be found that satisfies the tolerances specified in the optimization problem by adjusting the order of polynomial used in the approximation.

2.2. Interpolation and Numerical Integration

Interpolation and numerical integration serve an important role in the methods outlined in this dissertation and will be discussed in this section. The general structure of this section follows that provided by Giraldo in Chapters 4 and 5 of [44].

2.2.1. Interpolation

Polynomial interpolation is the method used to construct an N-th order polynomial, or interpolant, $x^N(t)$, that approximates a function, $x(t)$. This is typically done by ensuring the interpolant passes through the $N+1$ known points, $\{(t_i, x_i)\}_{i=0}^N$, so that $x(t_i) = x^N(t_i)$, for $i = 0, 1, \ldots, N$. This may be accomplished by using a finite sum such as

$$x^N(t) = \sum_{j=0}^{N} \Phi_j(t)\tilde{x}_j, \tag{2.24}$$

where $\{\tilde{x}_j\}_{j=0}^N$ are the expansion coefficients and $\{\Phi_j\}_{j=0}^N$ are the basis functions. Defining the basis functions, $\{\Phi_j\}_{j=0}^N$, as modes (such as Legendre polynomials) leads to *modal* type of interpolation. However, defining the basis functions in a nodal fashion such that $\Phi_j(t_i) = \delta_{ij}$, for $i, j = 0, 1, \ldots, N$, where

$$\delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases}$$

(such as Lagrange polynomials) leads to *nodal* interpolation.

2.2.1.1. Modal Interpolation

In modal interpolation, the basis functions, $\{\Phi_j\}_{j=0}^N$, in Equation (2.24) are typically orthogonal polynomials and the eigenfunctions of the singular Strurm-Liouville problem. Commonly used polynomials are: Legendre, Chebyshev, Fourier and Jacobi. For this discussion the Legendre polynomial, $L(t)$, will be the focus, defined by [45]

$$L_j(t) = \frac{(-1)^j}{2^j j!} \frac{d^j}{dt^j} \left((1-t^2)^j \right), \tag{2.25}$$

and therefore will be the chosen basis. The Legendre polynomials result from the special case of the singular Strurm-Liouville problem [45],

$$\frac{d}{dt} \left((1-t^2) \frac{dL_j(t)}{dt} \right) + j(j+1)L_j(t) = 0.$$

Figure 1 shows the first seven Legendre polynomials. The spectral coefficients, $\{a_j\}_{j=0}^{\infty}$,

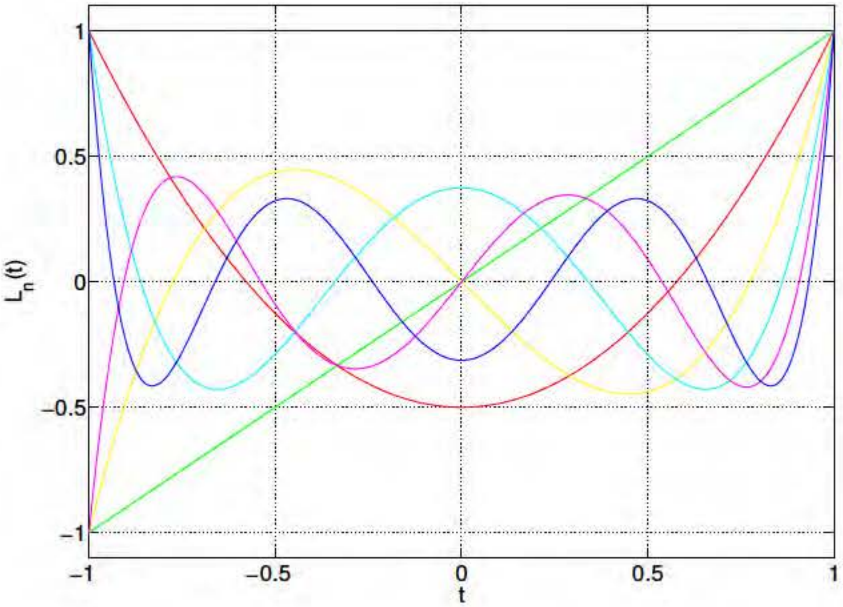


Figure 1: Legendre polynomials, $\{L_n(t)\}_{n=0}^{6}$.

for the continuous Legendre expansion, are defined as [46]

$$a_j = \frac{1}{\gamma_j} \int_{-1}^{1} x(t) L_j(t) dt, \tag{2.26}$$

where the normalizing constants, $\{\gamma_j\}_{j=0}^{\infty}$, for the Legendre polynomials are given by

$$\gamma_j = \frac{2}{2j+1}. \tag{2.27}$$

The truncated Legendre modal expansion is then given by

$$x^N(t) = \sum_{j=0}^{N} L_j(t) a_j. \tag{2.28}$$

However, due to the interpolatory nature of x^N, it is natural to seek spectral coefficients, $\{a_j\}_{j=0}^{N}$, defined by

$$x^N(t_i) = \sum_{j=0}^{N} L_j(t_i) a_j = \sum_{j=0}^{N} V_{ij} a_j, \tag{2.29}$$

for the known points $\{t_i\}_{i=0}^{N}$, where V is the generalized Vandermonde matrix given by [47]

$$V = \begin{pmatrix} L_0(t_0) & L_1(t_0) & \cdots & L_N(t_0) \\ L_0(t_1) & L_1(t_1) & \cdots & L_N(t_1) \\ \vdots & \vdots & \ddots & \vdots \\ L_0(t_N) & L_1(t_N) & \cdots & L_N(t_N) \end{pmatrix}. \tag{2.30}$$

From Equation (2.29), the modes, $\{a_j\}_{j=0}^{N}$, and the nodes, $\{\bar{x}^{Nj}\}_{j=0}^{N}$, are related by the generalized Vandermonde matrix (2.30) by the relationships [47]

$$\bar{x}^{Nj} = \sum_{j=0}^{N} V_{ij} a_j$$

and

$$a_i = \sum_{j=0}^{N} V_{ij}^{-1} \bar{x}^{Nj}, \tag{2.31}$$

where $\bar{x}^{Nj} = x^N(t_j)$ for $j = 0, 1, \ldots, N$.

Remark 2.2. *Note that to form Equation (2.31) the Vandermonde matrix (2.30) must be invertible and therefore nonsingular (and more practically speaking, well-conditioned). We will see that the invertibility of the Vandermonde matrix is dependent upon the interpolation quality of grid, $\{t_i\}_{i=0}^{N}$ (see Section 2.2.1.4 for grid selection)* [47]*. Additionally, we take comfort in the fact that the set of Legendre polynomials, $\{L_i\}_{i=0}^{\infty}$, is an orthogonal system that has shown to produce well-conditioned Vandermonde matrices for carefully selected nodes (as compared with the ill-conditioned Vandermonde matrices of non-orthogonal systems such as the the power basis, $\{t^n\}_{n=0}^{\infty}$)* [48].

Note that Equation (2.28) is a sum of frequencies, $\{L_j\}_{j=0}^{N}$, and amplitudes, $\{a_j\}_{j=0}^{N}$, that together compose the $(N+1)$ modes of $x^N(t)$. It is thus fitting to describe this approach as *modal* interpolation.

2.2.1.2. Nodal Interpolation

In nodal interpolation, the basis functions, $\{\Phi_j\}_{j=0}^{N}$, in Equation (2.24) are the Lagrange polynomials, $\{\phi_j^N\}_{j=0}^{N}$, of order N, defined on grid $\{t_i\}_{i=0}^{N}$, obtained from the general definition [45]

$$\phi_j^N(t) = \prod_{\substack{i=0 \\ i\neq j}}^{N} \frac{(t-t_i)}{(t_j-t_i)}. \tag{2.32}$$

Additionally, the Lagrange polynomials may be defined in terms of the Legendre polynomial by [46]

$$\phi_j^N(t) = \frac{1}{N(N+1)} \frac{(t^2-1)\,\dot{L}_N(t)}{(t-t_j)\,L_N(t_j)}. \tag{2.33}$$

Figure 2 shows the order $N = 6$ Lagrange polynomials, $\{\phi_j^N\}_{j=0}^{N}$, defined on an equispaced grid, $t \in [-1, 1]$.

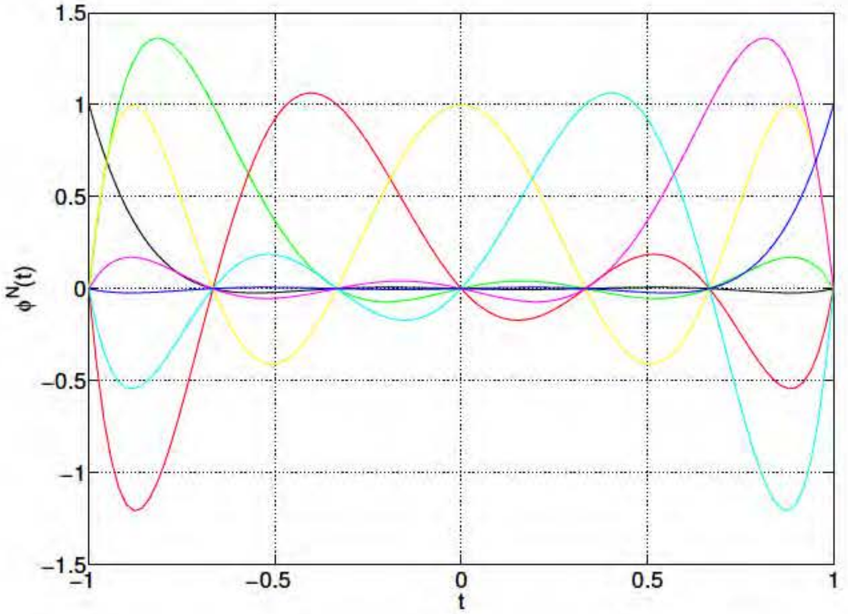


Figure 2: Lagrange polynomials of order $N = 6$ defined on an equi-spaced grid.

The nodal interpolation of the function $x(t)$ can be accomplished by the N-th order expansion

$$x^N(t) = \sum_{j=0}^{N} \phi_j^N(t) \bar{x}^{Nj}, \tag{2.34}$$

where $\bar{x}^{Nj} = x^N(t_j)$, for $j = 0, 1, \ldots, N$, since the Lagrange polynomial has the property, $\phi_j^N(t_i) = \delta_{ij}$.

Additionally, the Legendre polynomial, $L_i(t)$, of order i can be written as linear combinations of Lagrange polynomials, $\{\phi_i^N(t)\}_{i=0}^N$, of order N defined on grid, $\{t_i\}_{i=0}^N$, by the relationship [47]

$$L_i(t) = \sum_{j=0}^{N} L_i(t_j) \phi_j^N(t) = \sum_{j=0}^{N} V_{ij}^T \phi_j^N(t). \tag{2.35}$$

Likewise, the Lagrange polynomial may be written as linear combinations of Legendre polynomials by the relationship

$$\phi_i^N(t) = \sum_{j=0}^{N} \left(V^T\right)_{ij}^{-1} L_j(t) = \sum_{j=0}^{N} V_{ji}^{-1} L_j(t). \tag{2.36}$$

From Equations (2.31) and (2.36) we can relate the modal and nodal forms of the interplant, $x^N(t)$, by

$$\begin{aligned} x^N(t) =& \sum_{i=0}^{N} \phi_i^N(t)\bar{x}^{Ni} = \sum_{i=0}^{N} \left(\sum_{j=0}^{N} V_{ji}^{-1} L_j(t) \right) \bar{x}^{Ni} \\ =& \sum_{j=0}^{N} L_j(t) \left(\sum_{i=0}^{N} V_{ji}^{-1} \bar{x}^{Ni} \right) = \sum_{j=0}^{N} L_j(t) a_j. \end{aligned}$$

Therefore, Equation (2.34) is truly a *nodal* representation of Equation (2.28).

2.2.1.3. Transformations between grids

Consider the problem of transforming between two different grids, $\{t_j\}_{j=0}^{N}$ and $\{\tau_j\}_{j=0}^{M}$, where $M < N$. Let $\{\phi_j^M\}_{j=0}^{M}$ be the set of Lagrange polynomials of order M defined on grid $\{\tau_j\}_{j=0}^{M}$. Also, let the function $x^M(t)$ be the Lagrange interpolating polynomial

$$x^M(t) = \sum_{j=0}^{M} \phi_j^M(t)\bar{x}^{Mj},$$

where $\bar{x}^{Mj} = x^M(\tau_j)$, for $j = 0, 1, \ldots, N$, since $\phi_j^M(\tau_i) = \delta_{ij}$. Then the approximation of x^M at the dense gridpoints, $\{t_k\}_{k=0}^{N}$, can be calculated with the linear transformation

$$x^M(t_i) = \sum_{j=0}^{M} \phi_j^M(t_i)\bar{x}^{Mj} = \sum_{j=0}^{M} T_{ij}^{NM} \bar{x}^{Nj}, \quad i = 0, 1, \ldots, N,$$

where the $(N+1) \times (M+1)$ linear mapping, T^{NM}, is given by

$$T^{NM} = \begin{pmatrix} \phi_0^M(t_0) & \phi_1^M(t_0) & \cdots & \phi_M^M(t_0) \\ \phi_0^M(t_1) & \phi_1^M(t_1) & \cdots & \phi_M^M(t_1) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_0^M(t_N) & \phi_1^M(t_N) & \cdots & \phi_M^M(t_N) \end{pmatrix}. \tag{2.37}$$

In a similar fashion, the approximation of $\dot{x}$ may be transformed between two grids. Note that the approximation of $\dot{x}$ on grid $\{\tau_j\}_{j=0}^M$ may be given by

$$\dot{x}(t) \approx \dot{x}^M(t) = \sum_{j=0}^{M} \dot{\phi}_j^M(t)\bar{x}^{Mj},$$

where the derivative of the Lagrange polynomial is defined as

$$\dot{\phi}_j^M(t) = \sum_{\substack{k=0 \\ k\neq j}}^{M} \left(\frac{1}{t_j - t_k}\right) \prod_{\substack{i=0 \\ i\neq j \\ i\neq k}}^{M} \frac{t - t_i}{t_j - t_i}. \tag{2.38}$$

Then the approximation of $\dot{x}^M$ at the dense gridpoints, $\{t_j\}_{j=0}^N$, can be calculated with the with the linear transformation

$$\dot{x}^M(t_i) = \sum_{j=0}^{M} \dot{\phi}_j^M(t_i)\bar{x}^{Mj} = \sum_{j=0}^{M} A_{ij}^{NM}\bar{x}^{Mj}, \quad i = 0, 1, \ldots, N,$$

where the $(N+1) \times (M+1)$ linear mapping, A^{NM}, is given by

$$A^{NM} = \begin{pmatrix} \dot{\phi}_0^M(t_0) & \dot{\phi}_1^M(t_0) & \cdots & \dot{\phi}_M^M(t_0) \\ \dot{\phi}_0^M(t_1) & \dot{\phi}_1^M(t_1) & \cdots & \dot{\phi}_M^M(t_1) \\ \vdots & \vdots & \ddots & \vdots \\ \dot{\phi}_0^M(t_N) & \dot{\phi}_1^M(t_N) & \cdots & \dot{\phi}_M^M(t_N) \end{pmatrix}. \tag{2.39}$$

Using the transformation matrices (2.37) and (2.39) to relate different grids will serve as an important tool for the multi-scale approximation methods outlined in Chapters 3 and 5. However, the accuracy of interpolation is extremely important and will be discussed next.

2.2.1.4. Interpolation Quality

Approximation quality is a great concern when using interpolation. The goodness of the approximation $x^N(t)$ is directly related to the grid points, $\{t_j\}_{j=0}^N$, from which the Lagrange polynomials, $\{\phi_j^N\}_{j=0}^N$, are defined. A measure of interpolation goodness is the Lebesque constant, Λ_N, given by [45]

$$\Lambda_N = \max_{t\in[-1,1]} \sum_{j=0}^{N} \left|\phi_j^N(t)\right|. \tag{2.40}$$

The best interpolating polynomial $x^N(t)$ is one that minimizes the Lebesgue constant (2.40), due to the following result [45],

$$\left\|x(t) - x^N(t)\right\|_{L^\infty} \leq (1 + \Lambda_N)\|x(t) - p(t)\|_{L^\infty},$$

where $p(t)$ is the best approximating polynomial of $x(t)$ in the L^∞-norm (see Appendix A). From [45, 49], for any set of $(N + 1)$ distinct points, $t_i \in [-1, 1]$, for $i = 0, 1, \ldots, N$, the Lesbegue constant (2.40) has the lower bound [45],

$$\frac{2}{\pi}\log(N + 1) + \alpha \leq \Lambda_N,$$

where $\alpha = \frac{2}{\pi}\left(\gamma + \log\frac{4}{\pi}\right) \approx 0.521$ and $\gamma = 0.57721566...$ is the Euler-Mascheroni constant. So, at best the selected grid is associated with a Lebesgue constant that grows logarithmically [45]. Common Legendre family of points used for interpolation are the Legendre-Gauss (LG), Legendre-Gauss-Lobatto (LGL) and Legendre-Gauss-Radau (LGR) points.

Legendre-Gauss Points. The LG points, $\{t_i\}_{i=0}^{N}$, are defined by $-1 < t_0 < \cdots < t_N < 1$, and are the roots of

$$\xi(t) = L_{N+1}(t), \tag{2.41}$$

where $L_{N+1}(t)$ is the $(N+1)$-th order Legendre polynomial. Note that the LG points do not include the endpoints, $t = \pm 1$. Figure 3 shows the LG points for various orders of N.

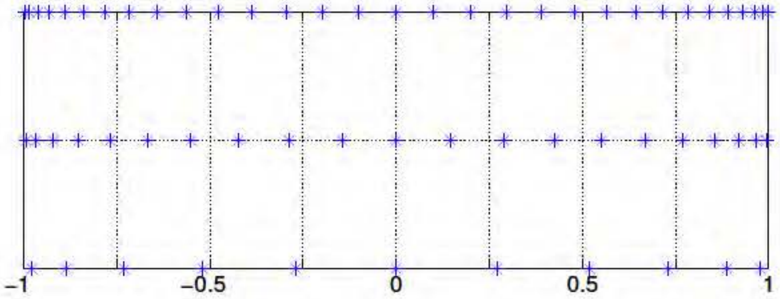


Figure 3: LG points for $N = 10, 20$ and 30.

Legendre-Gauss-Lobatto Points. The LGL points, $\{t_i\}_{i=0}^{N}$, are defined by $t_0 = -1 < t_1 < \cdots < t_N = 1$, and are the roots of

$$\xi(t) = (1 - t^2)\dot{L}_N(t), \tag{2.42}$$

where $\dot{L}_N(t)$ is the derivative of the N-th order Legendre polynomial. Note that the LGL points include the endpoints, $t = \pm 1$. Figure 4 shows the LG points for various orders of N.

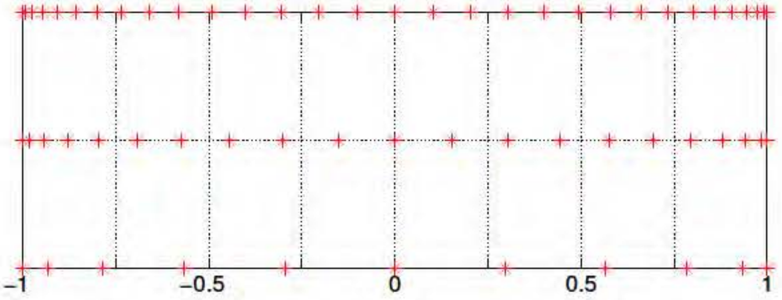


Figure 4: LGL points for $N = 10, 20$ and 30.

Additionally, Figure 5 shows the order $N = 6$ Lagrange polynomials, ϕ^N, defined on a LGL grid.

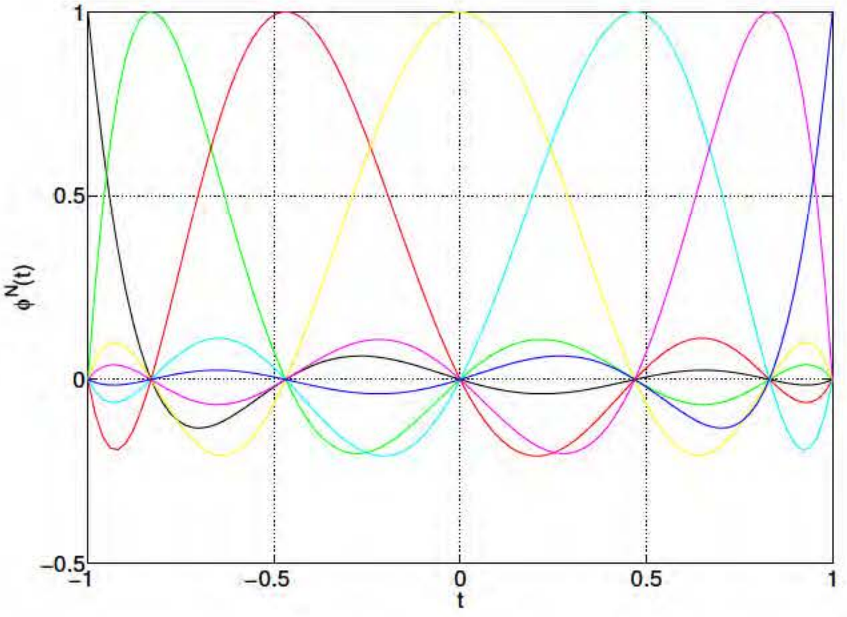


Figure 5: Lagrange polynomials of order $N = 6$ defined on a LGL grid.

Legendre-Gauss-Radau Points. The LGR points, $\{t_i\}_{i=0}^N$, are defined by $t_0 = -1 < t_1 < \cdots < t_N < 1$, and are the roots of

$$\xi(t) = L_{N+1}(t) + L_N(t). \tag{2.43}$$

Note that the LGR points only include the endpoint, $t = -1$. Figure 6 shows the LGR points for various orders of N.

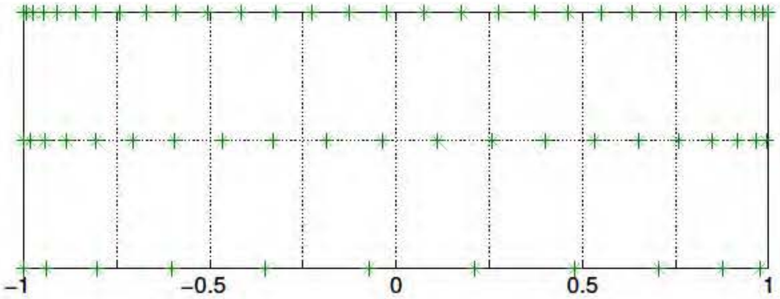


Figure 6: LGR points for $N = 10, 20$ and 30.

Additionally, flipped-LGR (F-LGR) points are the negative of the LGR points and are therefore defined by $-1 < t_0 < \cdots < t_N = 1$. Note that the F-LGR points only include the endpoint, $t = 1$.

Although all three sets of Legendre points (LG, LGL and LGR) have Lebesgue constants (2.40) that grow logarithmically or sublinearly with N, the LGL grid is asymptotically associated with the near optimal Lebesgue constant [45],

$$\Lambda_N^{LGL} \leq \frac{2}{\pi}\log(N+1) + 0.685...$$

As alluded to in Remark 2.2, the quality of interpolation can also be observed by analyzing the conditioning of the Vandermonde matrix (2.30). Due to the relationship between the Legendre polynomials, $\{L_j\}_{j=0}^N$, and Lagrange polynomials, $\{\phi_j^N\}_{j=0}^N$, shown in (2.35), Cramer's rule [50] provides the following relationship

$$\phi_j^N = \frac{\text{Det}\left[\mathbf{L}(t_0), \ldots, \mathbf{L}(t_{j-1}), \mathbf{L}(t), \mathbf{L}(t_{j+1}), \ldots, \mathbf{L}(t_N)\right]}{\text{Det}\left[V^T\right]}, \tag{2.44}$$

where $\mathbf{L}(t) = [L_0(t), L_1(t), \ldots, L_N(t)]^T$. As pointed out by Hesthaven et al. [47], if the goal is to minimize the Lebesque constant (2.40), we should strive to maximize the denominator of Equation (2.44), $\text{Det}\big[V^T\big]$. This leads to the LGL grid set [51]. Additionally, the Chebyshev-Gauss family of points proves to have excellent interpolation quality, particularly the Chebyshev-Gauss-Lobatto points when measured by the Lebesgue constant growth [45, 52]. However, the focus in this dissertation will be on the Legendre basis, and thus the LG, LGL and LGR points.

Unfortunately, equi-spaced points prove to be a very poor grid selection for interpolation. The Lebesgue constant for the equi-spaced points grows asymptotically like [52, 53],

$$\Lambda_N^{ES} \sim \frac{2^{N+1}}{eN(\log N + \gamma)},$$

very far from optimal.

As an example of interpolation quality consider the function

$$f(t) = \cos(\mu \pi t), \quad t \in [-1, 1], \tag{2.45}$$

with $\mu = 3$, shown in Figure 7.

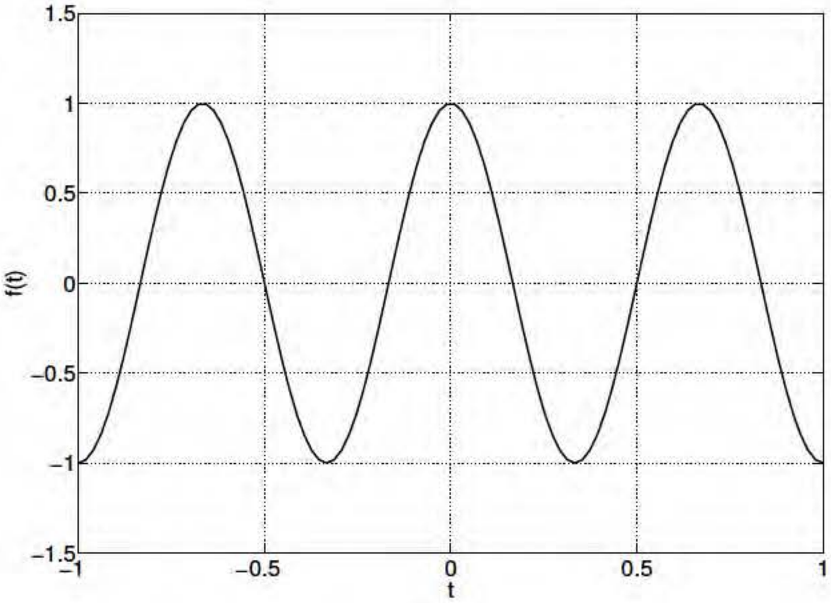


Figure 7: Plot of $f(t) = \cos(3\pi t)$.

When low order approximations are used to interpolate $f(t)$, inaccuracies are apparent. Figure 8 shows the inaccuracies in the 10-th order Lagrange interpolating polynomial approximation of $f(t)$ with LGL points.

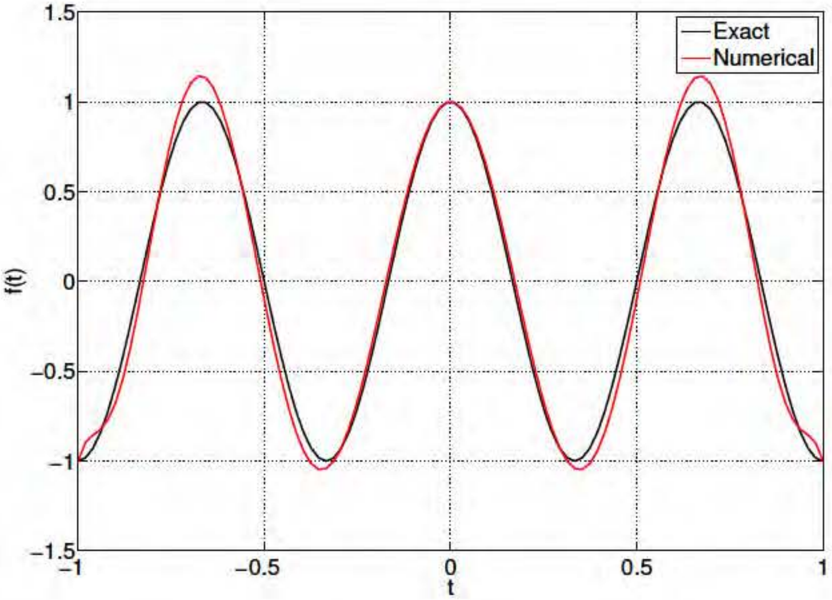


Figure 8: Interpolation of $f(t) = \cos(3\pi t)$ with 10-th order LGL points.

However, as the interpolation order, N, is increased, the maximum error,

$$\|\text{error}\|_{\infty} = \left\| f(t_i) - f^N(t_i) \right\|_{\infty}, \quad i = 0, 1, \ldots, N,$$

decreases exponentially with N, where $\| \zeta \|_{\infty}$ represents the maximum element of vector, $\zeta \in \mathbb{R}^n$. Figure 9 shows the visual accuracy of the 30-th order Lagrange interpolating polynomial of $f(t)$ with LGL points.

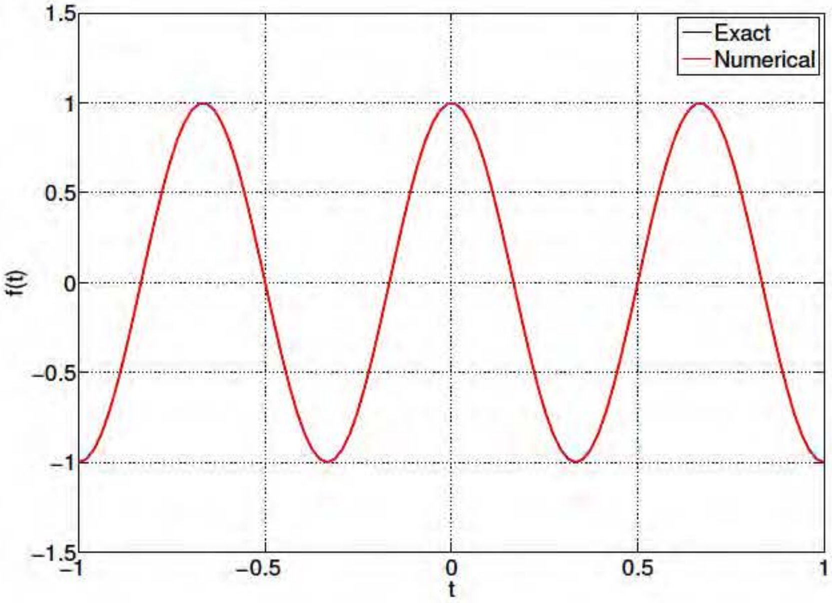


Figure 9: Interpolation of $f(t) = \cos(3\pi t)$ with 30-th order LGL points.

Additionally, Figure 10 compares interpolation of $f(t)$ with equi-spaced, LG, LGL and LGR points, for various orders of N. Notice that for LG, LGL and LGR points, the maximum interpolation error drops to $O(10^{-15})$ by $N = 35$. However, in general, the equi-spaced points prove to have very poor interpolation quality.

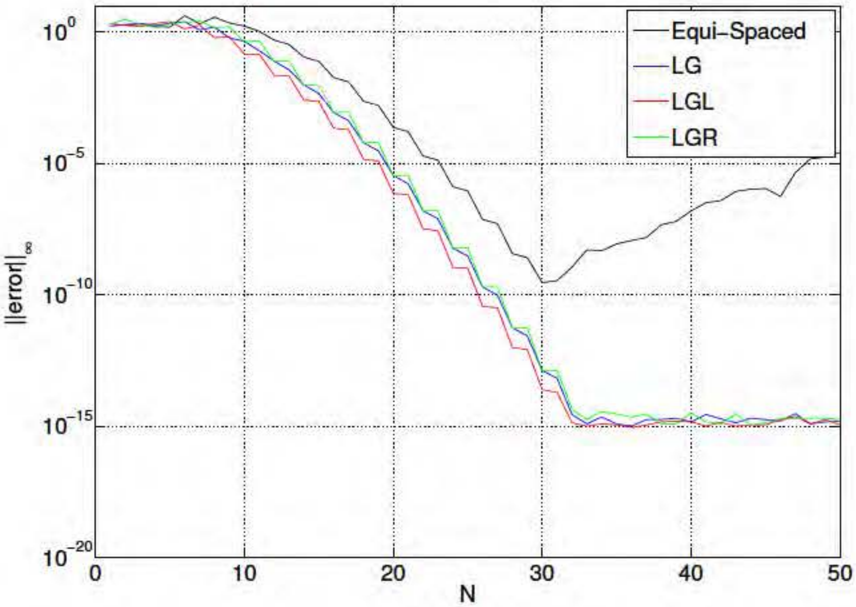


Figure 10: Comparison of interpolation errors for various orders of N and equi-spaced, LG, LGL and LGR points.

Due to the accuracy of LGL interpolation, and inclusion of the endpoints, $t = \pm 1$, LGL points are readily used for numerical computation. However, also related to point selection is the accuracy of numerical integration. This is an important factor for the direct methods for optimal control (due to the cost function that is normally approximated by numerical integration) and will be discussed next.

2.2.2. Numerical Integration

Numerical integration, or quadrature, is a way of approximating an integral with a sum

$$\int_{-1}^{1} x(t)dt \approx \sum_{k=0}^{N} x(t_k)w_k,$$

where $\{w_k\}_{k=0}^N$ are the quadrature weights and $\{t_k\}_{k=0}^N$ are the associated points. Ideally, the numerical integration is exact, but it is reasonable to expect an error, ϵ^N, such that

$$\epsilon^N = \int_{-1}^{1} x(t)dt - \sum_{k=0}^{N} x(t_k)w_k.$$

For the special case of the numerical integration of general function $x \in C^{N+1}$, that is approximated by the Lagrange interpolating polynomial

$$x(t) \approx x^N(t) = \sum_{j=0}^{N} \phi_j^N(t)\bar{x}^{Nj},$$

the error may be given by [54]

$$\epsilon^N = \frac{1}{(N+1)!}\int_{-1}^{1}\prod_{j=0}^{N}(t-t_j)\frac{d^{(N+1)}x(\xi(t))}{dt^{(N+1)}}dt,$$

for arbitrary function $\xi(t) \in [-1,1]$.

However, for certain classes of polynomial functions, the quadrature error is zero. Consider the function $x(t) \in P_N$, represented as the finite sum

$$x(t) = \sum_{j=0}^{N} \phi_j^N(t)\bar{x}^{Nj}.$$

Performing numerical integration on x results in

$$\begin{aligned} \int_{-1}^{1} x(t)dt &= \int_{-1}^{1}\sum_{j=0}^{N}\phi_j^N(t)\bar{x}^{Nj}dt = \sum_{j=0}^{N}\bar{x}^{Nj}\int_{-1}^{1}\phi_j^N(t)dt \\ &= \sum_{j=0}^{N}\bar{x}^{Nj}\sum_{k=0}^{N}\phi_j^N(t_k)w_k = \sum_{j=0}^{N}\bar{x}^{Nj}w_j. \end{aligned}$$

Clearly, since

$$\int_{-1}^{1} x(t)dt = \sum_{j=0}^{N} \bar{x}^{Nj} w_j, \tag{2.46}$$

quadrature is exact $\forall x(t) \in P_N$. Additionally, the quadrature weights, $\{w_j\}_{j=0}^{N}$, can be found with the relationship

$$w_j = \int_{-1}^{1} \phi_j^N(t)dt. \tag{2.47}$$

2.2.2.1. Gaussian Quadrature

Consider now the case that $x(t) \in P_{2N+1}$ written in the form

$$x(t) = L_{N+1}(t)f(t) + g(t),$$

where $f, g \in P_N$, L_{N+1} is the Legendre polynomial of order $(N+1)$ and P_N denotes the space of all polynomials of degree $\leq N$. Also consider the $(N+1)$ points, $\{t_k\}_{k=0}^{N}$, that are the roots of $L_{N+1}(t)$ (known as LG points, discussed in Section 2.2.1), and the associated quadrature weights, $\{w_k\}_{k=0}^{N}$ (known as LG quadrature weights, found via the general definition (2.47) or the more specific definition (2.48)). Then $x(t_k) = g(t_k)$, for all $k = 0, 1, \ldots, N$, and

$$\begin{aligned} \int_{-1}^{1} x(t)dt &= \int_{-1}^{1} \left(L_{N+1}(t)f(t) + g(t)\right) dt \\ &= \sum_{k=0}^{N} L_{N+1}(t_k)f(t_k)w_k + \sum_{k=0}^{N} g(t_k)w_k \\ &= \sum_{k=0}^{N} g(t_k)w_k = \sum_{k=0}^{N} x(t_k)w_k. \end{aligned}$$

This is known as LG quadrature (or simply Gauss quadrature), which is exact $\forall x(t) \in P_{2N+1}$. In the case that the function $x(t)$ is a polynomial, such that $x(t) \in P_{2N+\delta}$, numerical integration is exact for LGL and LGR quadrature, where $\delta = -1 \text{ and } 0$, respectively. The proof of LGL and LGR quadrature exactness for polynomials is similar to that given above for LG quadrature. The list of LG, LGL and LGR weights are presented below (provided by [47]) and point locations are given by Equations (2.41), (2.42) and (2.43), respectively.

Legendre-Gauss Quadrature. The LG quadrature weights, $\{w_k\}_{k=0}^{N}$, are given by

$$w_k = \frac{2}{[1-(t_k)^2][\dot{L}_{N+1}(t_k)]^2}. \tag{2.48}$$

Legendre-Gauss-Lobatto Quadrature. The LGL quadrature weights, $\{w_k\}_{k=0}^{N}$, are given by

$$w_k = \frac{2}{N(N+1)} \frac{1}{[L_N(t_k)]^2}. \tag{2.49}$$

Legendre-Gauss-Radau Quadrature. The LGR quadrature weights, $\{w_k\}_{k=0}^{N}$, are given by

$$w_k = \frac{1}{(N+1)^2} \frac{1-t_k}{[L_N(t_k)]^2}. \tag{2.50}$$

Additionally, the F-LGR quadrature weights, $\{\tilde{w}_k\}_{k=0}^{N}$, are the reordered LGR quadrature weights, $\{\tilde{w}_k\}_{k=0}^{N} = \{w_{N-k}\}_{k=0}^{N}$.

2.2.2.2. Gaussian Quadrature Accuracy

The Legendre-Gaussian family of quadrature is widely used due to its integrating accuracy and has the following error estimates for LG, LGL and LGR quadrature, respectively [55]

$$\epsilon_{LG}^{N} = \frac{2^{2N+3}[(N+1)!]^4}{(2N+3)[(2N+2)!]^3}\frac{d^{(2N+2)}x(\xi)}{dt^{(2N+2)}},\ \xi \in [-1,1],$$

$$\epsilon_{LGL}^{N} = \frac{-(N+1)N^3 2^{2N+1}[(N-1)!]^4}{(2N+1)[(2N)!]^3}\frac{d^{(2N)}x(\xi)}{dt^{(2N)}},\ \xi \in [-1,1],$$

$$\epsilon_{LGR}^{N} = \frac{2^{2N+1}(N+1)(N!)^4}{[(2N+1)!]^3}\frac{d^{(2N+1)}x(\xi)}{dt^{(2N+1)}},\ \xi \in [-1,1].$$

Consider again the function (2.45), with $\mu = 1/2$. As an example of quadrature accuracy, consider the numerical approximation,

$$\int_{-1}^{1} \cos\left(\frac{\pi t}{2}\right) dt \approx \sum_{k=0}^{N} f^N(t_k) w_k, \tag{2.51}$$

where

$$f(t) \approx f^N(t) = \sum_{j=0}^{N} \phi_j^N(t) \bar{f}^{Nj},$$

and $\bar{f}^{Nj} = \cos(\frac{\pi t_j}{2})$, for $j = 0, 1, \ldots, N$. Due to the property of the Lagrange polynomial, $\phi_j^N(t_i) = \delta_{ij}$, the quadrature expression in (2.51) takes the form (2.46),

$$\sum_{k=0}^{N} f^N(t_k) w_k = \sum_{k=0}^{N} \bar{f}^{Nk} w_k.$$

As with interpolation error, quadrature error decreases exponentially with N. Figure 11 shows numerical integration error,

$$\text{error} = \left| \int_{-1}^{1} \cos(\frac{\pi t}{2}) dt - \sum_{k=0}^{N} \cos(\frac{\pi t_k}{2}) w_k \right|,$$

for LG, LGL and LGR quadrature, all of which demonstrate similar exponential convergence as N increases.

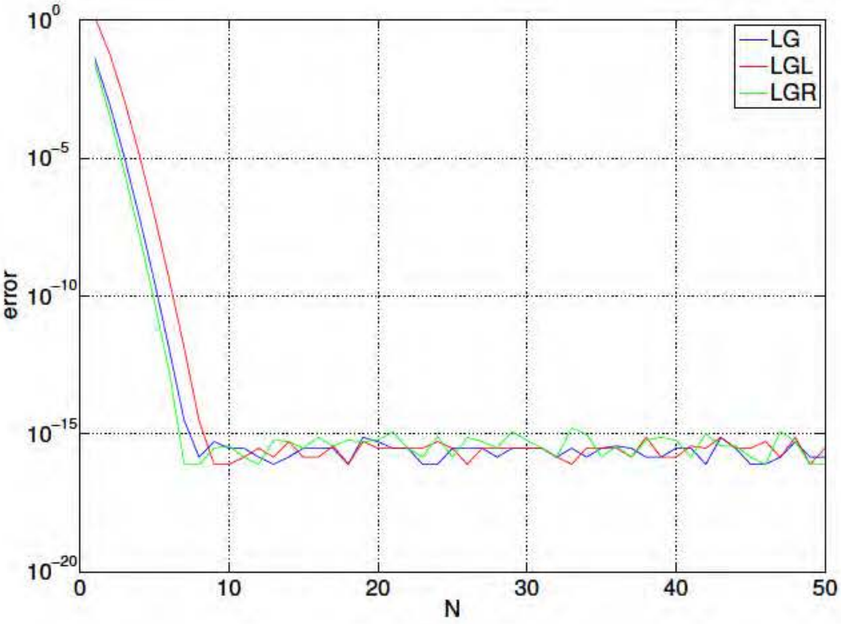


Figure 11: Comparison of quadrature errors for various orders of N and LG, LGL and LGR points.

Due to the accuracy of quadrature, quality of interpolation and inclusion of the endpoints, $t = \pm 1$, LGL points are an important part of many numerical computation applications, to include direct methods for optimal control. For instance, in the Legendre PS method the problem state-control constraints are discretized using LGL points. Additionally, the states are approximated with globally interpolating Lagrange polynomials defined on an LGL grid and the cost function is approximated using LGL quadrature rule. LGL interpolation and quadrature also serve an important role in the construction of the Galerkin optimal control formulations discussed in this dissertation.

A crucial step in solving optimal control problems with direct methods—and yet to be discussed—is the discretization of the system dynamics. Although a number of techniques have been investigated to do this, spectral methods have proven to be effective and efficient. This is a highlight in the PS direct methods for optimal control, where collocation methods are used to discretize system dynamics. Spectral methods also serve as the heart of Galerkin optimal control, where Galerkin methods are employed. Both collocation and Galerkin methods will be discussed in the next section.

2.3. Numerical Solutions to Differential Equations

Spectral methods, which have gained much popularity due to their spectral accuracy and versatility [56, 57], can be formulated for both local (element-based) and global approximations. Consider the task of discretizing the dynamics of problem (2.5)–(2.8),

$$\dot{x}(t) = f(x(t), u(t)),\ t \in [t_0, t_f], \tag{2.52}$$

where $t \mapsto (x, u) \in \mathbb{R}^{N_x} \times \mathbb{R}^{N_u}$, with initial conditions,

$$x(t_0) = x_0, \tag{2.53}$$

and endpoint conditions,

$$e(x(t_f)) = 0, \tag{2.54}$$

where $f : \mathbb{R}^{N_x} \times \mathbb{R}^{N_u} \to \mathbb{R}^{N_x}$ and $e : \mathbb{R}^{N_x} \times \mathbb{R}^{N_x} \to \mathbb{R}^{N_e}$, are Lipschitz continuous with respect to their argument. This can be accomplished using a number of spectral method formulations. However, two methods will be discussed here, collocation and Galerkin. Additionally, of the spectral element methods, continuous and discontinuous Galerkin element-based formulations will be outlined.

2.3.1. Spectral Methods

The starting point for the spectral approximation of Equation (2.52) is to approximate the solutions x and u by the finite sums

$$x(\xi) \approx x^N(\xi) = \sum_{j=0}^{N} \Phi_j(\xi)\hat{x}_j,$$
$$u(\xi) \approx u^N(\xi) = \sum_{j=0}^{N} \tilde{\Phi}_j(\xi)\hat{u}_j,$$

where $\hat{x}_j$ and $\hat{u}_j$ are expansion coefficients and Φ_j and $\tilde{\Phi}_j$ are basis functions. In terms of the approximation x^N and u^N, Equation (2.52) becomes

$$\dot{x}^N(\xi) - \frac{\Delta t}{2} f(x^N(\xi), u^N(\xi)) = \epsilon^N(\xi), \tag{2.55}$$

where $\Delta t = t_f - t_0$ and ϵ^N is the error (or residual) in the approximation which, generally, is not zero. The relationship between the physical time domain, $t \in [t_0, t_f]$, and the computational space, $\xi \in [-1, 1]$, is given by

$$\xi = \frac{2}{\Delta t}(t - t_0) - 1 \quad \text{and} \quad d\xi = \frac{2}{\Delta t} dt,$$

and conversely,

$$t = \frac{\Delta t}{2}(\xi + 1) + t_0 \quad \text{and} \quad dt = \frac{\Delta t}{2} d\xi.$$

2.3.1.1. Collocation

In the collocation method the basis functions, $\{\Phi_j\}_{j=0}^N$ are the Lagrange polynomials (2.32), $\{\phi_j^N\}_{j=0}^N$, of order N, defined on the grid of collocation points, $\{\xi_j\}_{j=0}^N \in [-1, 1]$; while $\{\tilde{\Phi}_j\}_{j=0}^N = \{\psi_j^N\}_{j=0}^N$, where $\{\psi_j^N\}_{j=0}^N$ is any continuous function (not necessarily a polynomial) with the property $\psi_j(\xi_i) = \delta_{ij}$. The expansion coefficients are, $\hat{x}_j = \bar{x}^{Nj}$ and $\hat{u}_j = \bar{u}^{Nj}$, therefore $x^N(\xi_j) = \bar{x}^{Nj}$ and $u^N(\xi_j) = \bar{u}^{Nj}$, for $j = 0, 1, \ldots, N$.

The approximations of x and u in the computational space, ξ, are given by

$$x^N(\xi) = \sum_{j=0}^{N} \phi_j^N(\xi)\bar{x}^{Nj},$$
$$u^N(\xi) = \sum_{j=0}^{N} \psi_j^N(\xi)\bar{u}^{Nj}.$$

The approximation of $\dot{x}$ in the computational space, ξ, is then

$$\dot{x}^N(\xi) = \sum_{j=0}^{N} \dot{\phi}_j^N(\xi)\bar{x}^{Nj},$$

where the derivative of the Lagrange polynomial, $\{\dot{\phi}_j^N\}_{j=0}^N$, is given by Equation (2.38). Additionally, in the collocation method, the error term, ϵ^N, is ideally forced to zero at each collocation point, therefore, Equation (2.55) becomes

$$\sum_{j=0}^{N} \dot{\phi}_j^N(\xi_i)\bar{x}^{Nj} - \frac{\Delta t}{2} f(\bar{x}^{Ni}, \bar{u}^{Ni}) = 0, \quad i = 0, 1, \ldots, N.$$

Collocation methods assume the title *pseudospectral* methods, due to the nodal nature of the formulation—in lieu of *spectral* referring to a transformation from physical to spectral space. Common Legendre family of collocation nodes used are the LGL, LG and LGR points. When LGL nodes, $\{\xi_i\}_{i=0}^N$, defined by $-1 = \xi_0, \xi_1, \ldots, \xi_{N-1}, \xi_N = 1$, are used for the discretization, the Legendre PS differentiation matrix, A, is given by $A_{ij} = \dot{\phi}_j^N(\xi_i)$ for $i, j = 0, 1, \ldots, N$, resulting in the system of equations,

$$\sum_{j=0}^{N} A_{ij}\bar{x}^{Nj} - \frac{\Delta t}{2} f(\bar{x}^{Ni}, \bar{u}^{Ni}) = 0, \quad i = 0, 1, \ldots, N. \tag{2.56}$$

In addition to Equation (2.38), the Legendre PS differentiation matrix, A, may be defined in terms of the Legendre polynomials by [46]

$$A_{ij} = \begin{cases} \frac{L_N(\xi_i)}{L_N(\xi_j)} \frac{1}{\xi_i - \xi_j} & i \neq j, \\ -\frac{N(N+1)}{4} & i = j = 0, \\ \frac{N(N+1)}{4} & i = j = N, \\ 0 & i = j \epsilon \left[1, \ldots, N-1\right]. \end{cases} \tag{2.57}$$

Remark 2.3. *The derivative of $x^N(\xi)$ at each LGL point $\{\xi_i\}_{i=0}^{N}$ is exactly equal to*

$$\dot{x}^N(\xi_i) = \sum_{j=0}^{N} A_{ij} \bar{x}^{Nj},$$

for any polynomial with degree less than or equal to N [46]*. However, a feasible solution to the equality dynamical constraint may not exist. In order to guarantee feasibility of the discretized problem, Gong et al.* [5]*, suggest a relaxation of the equality constraint.*

Therefore, Equation (2.52) may be discretized with the following inequality constraint,

$$\left\| \sum_{j=0}^{N} A_{ij} \bar{x}^{Nj} - \frac{\Delta t}{2} f(\bar{x}^{Ni}, \bar{u}^{Ni}) \right\|_{\infty} \leq \delta^N, \quad i = 0, 1, \ldots, N,$$

where δ^N is the feasibility tolerance that is dependent on N and the smoothness of x and u (see Section 3.2); and $\parallel \zeta \parallel_\infty$ represents the maximum element of vector, $\zeta \in \mathbb{R}^n$. The initial conditions and endpoint conditions may be approximated similarly by

$$\left\| \bar{x}^{N0} - x_0 \right\|_{\infty} \leq \delta^N \quad \text{and} \quad \left\| e(\bar{x}^{NN}) \right\|_{\infty} \leq \delta^N.$$

Collocation methods have become popular for the discretization of system dynamics in direct methods for optimal control, specifically psuedospectral methods. For the Legendre PS method, the LGL points become the discretization of choice.

2.3.1.2. Galerkin Numerical Methods

Galerkin methods can be subdivided into two main categories, Bubnov-Galerkin and Petrov-Galerkin. The weighted residual method forms the basis for the development of these approximation techniques and will help to distinguish them [57].

For the weighted residual method, the weak integral form of the Equation (2.55) is solved by multiplying by a test function, Ψ_i, integrating over the domain, and ideally we force the residual term to zero,

$$\int_{-1}^{1} \Psi_i(\xi) \left(\dot{x}^N(\xi) - \frac{\Delta t}{2} f(x^N(\xi), u^N(\xi)) \right) d\xi = \int_{-1}^{1} \Psi_i(\xi) \epsilon^N(t) d\xi = 0, \tag{2.58}$$

for $i = 0, 1, \ldots, N$.

Remark 2.4. *Setting the residual terms to zero in Equation (2.58), ($\int_{-1}^{1} \Psi_i(\xi)\epsilon^N(\xi)d\xi = 0$, for each $i = 0, 1, \ldots, N$), is akin to forcing the orthogonality of the space spanned by Ψ_i and ϵ^N in $L^2[-1, 1]$.*

The approximation x^N and u^N can be found satisfying Equation (2.58). Common test functions include orthogonal polynomials (such as the Legendre polynomials, L) and the trigonometric functions. In the Bubnov-Galerkin method, the test functions are the same as the basis functions, unlike the Petrov-Galerkin method, where the test and basis functions are different. The general structure of these global Galerkin methods—as well as the mathematical notation used in this dissertation—is provided by Giraldo [44] and discussed in the following sections.

Bubnov-Galerkin In the *Bubnov-Galerkin method* (or often called simply the *Galerkin method*) the test and basis functions are the same. These functions can be *modal* or *nodal* in nature, however, in this section the focus will be on *nodal* Galerkin methods. For a nodal Galerkin approach, it is common to use a Legendre based grid such as the LGL, LG or LGR nodes, and define the test and basis functions as Lagrange polynomials (2.32), $\{\phi_j^N\}_{j=0}^N$, of order N, on the selected grid. A popular selection for interpolation points are the LGL nodes due to the accuracy of LGL quadrature and the inclusion of endpoints, $t = \pm 1$. For this discussion, the LGL nodes will be the focus.

The approximations of x and u in the computational space, $\xi \in [-1, 1]$, are given by

$$x^N(\xi) = \sum_{j=0}^{N} \phi_j^N(\xi)\bar{x}^{Nj},$$
$$u^N(\xi) = \sum_{j=0}^{N} \psi_j^N(\xi)\bar{u}^{Nj}.$$

where $\{\psi_j^N\}_{j=0}^N$ is any continuous function (not necessarily a polynomial) with the property $\psi_j(\xi_i) = \delta_{ij}$. The expansion coefficients are, $\hat{x}_j = \bar{x}^{Nj}$ and $\hat{u}_j = \bar{u}^{Nj}$, therefore $x^N(\xi_j) = \bar{x}^{Nj}$ and $u^N(\xi_j) = \bar{u}^{Nj}$, for $j = 0, 1, \ldots, N$. Equation (2.58) becomes [44]

$$\int_{-1}^{1} \phi_i^N(t)\dot{\phi}_j^N(\xi)d\xi \, \bar{x}^{Nj} - \frac{\Delta t}{2}\int_{-1}^{1} \phi_i^N(\xi) f(x^N(\xi), u^N(\xi))d\xi = 0, \tag{2.59}$$

for $i = 0, 1, \ldots, N$, or using matrix-vector notation,

$$\sum_{j=0}^{N} D_{ij}\bar{x}^{Nj} - c_i = 0, \quad i = 0, 1, \ldots, N.$$

The Galerkin differentiation matrix, D, and RHS vector, c are defined as

$$D_{ij} = \int_{-1}^{1} \phi_i^N(\xi)\dot{\phi}_j^N(\xi)d\xi,$$
$$c_i = \frac{\Delta t}{2}\int_{-1}^{1} \phi_i^N(\xi) f(x^N(\xi), u^N(\xi))d\xi,$$

for $i, j = 0, 1, \ldots, N$.

Using LGL quadrature, D can be calculated with the relationship

$$D_{ij} = \sum_{k=0}^{Q} \phi_i^N(\xi_k)\dot{\phi}_j^N(\xi_k)w_k = \dot{\phi}_j^N(\xi_i)w_i = A_{ij}w_i, \quad i, j = 0, 1, \ldots, N,$$

where $\{w_i\}_{i=0}^N$ are the LGL weights given by Equation (2.49) and A is the Legendre PS differentiation matrix (2.57). Since LGL quadrature rule is exact for polynomial integrands of degree less than or equal to $2N-1$, the numerical integration is done exactly when $Q = N$ LGL integration points are used.

Using LGL quadrature rule, c can be approximated by the relationship

$$c_i \approx \frac{\Delta t}{2}\sum_{k=0}^{Q} \phi_i^N(\xi_k) f(x^N(\xi_k), u^N(\xi_k)) w_k, \quad i = 0, 1, \ldots, N.$$

When $Q = N$ LGL quadrature points are used, the RHS vector approximation, $\bar{c}^N$, can be expressed in the simplified form

$$\bar{c}^{Ni} = \frac{\Delta t}{2} f(\bar{x}^{Ni}, \bar{u}^{Ni}) w_i, \quad i = 0, 1, \ldots, N.$$

Remark 2.5. *Recall that for LGL quadrature rule, integration is exact for polynomial integrands of degree less than or equal to $2N-1$. If $Q = (N+1)$ integration points are used, the RHS vector will integrate exactly when $f(x(t), u(t))$ is linear in x and u. In the case of a nonlinear function f, the accuracy of integration (and therefore the accuracy of the overall approximation) can be improved by increasing the number of quadrature points Q.*

When $Q = N$ LGL quadrature points are used to calculate the Galerkin differentiation matrix and approximate the RHS vector, the system may be simplified as

$$\sum_{j=0}^{N} D_{ij}\bar{x}^{Nj} - \bar{c}^{Ni} = 0, \quad i = 0, 1, \ldots, N. \tag{2.60}$$

Remark 2.6. *In form (2.60), the resulting Galerkin equations that must be satisfied are*

$$\left(\sum_{j=0}^{N} A_{ij}\bar{x}^{Nk} - \frac{\Delta t}{2} f(\bar{x}^{Ni}, \bar{u}^{Ni})\right) w_i = 0, \tag{2.61}$$

for $i = 0, 1, \ldots, N$. Note that the relationship in parentheses,

$$\sum_{j=0}^{N} A_{ij}\bar{x}^{Nk} - \frac{\Delta t}{2} f(\bar{x}^{Ni}, \bar{u}^{Ni}) = 0,$$

for $i = 0, 1, \ldots, N$, are the same equations that would be satisfied when using the collocation method (see Equation [2.56]). For this reason, Bubnov Galerkin with numerical integration is sometimes called the "collocation method in the weak form." [57]

In the words of John Boyd, "collocation—with the right set of points—must inherit the aura of invincibility of the Galerkin method." [58]

Remark 2.7. *An inequality version of (2.61) has been known and used in pseudospectral optimal control methods. Details on its relationship with Galerkin optimal control are addressed in Chapter 4 in Remark 4.2.*

Due to the results of Gong et al. [5], we know a feasible solution to the equality dynamical constraint may not exist. In order to guarantee feasibility of the discretized problem, the following inequality constraint is suggested,

$$\left\| \sum_{j=0}^{N} D_{ij}\bar{x}^{Nj} - \bar{c}^{Ni} \right\|_{\infty} \leq \delta^N, \quad i = 0, 1, \ldots, N, \tag{2.62}$$

where δ^N is the feasibility tolerance that is dependent on N and the smoothness of x and u (see Chapter 4). The initial conditions and endpoint conditions may be approximated similarly by

$$\left\|\bar{x}^{N0} - x_0\right\|_{\infty} \leq \delta^N \quad \text{and} \quad \left\|e(\bar{x}^{NN})\right\|_{\infty} \leq \delta^N.$$

Remark 2.8. *The inequality formulation (2.62) introduces some fundamental differences in numerical analysis. In the Galerkin approach, the error is measured by the L^2-norm. As a result, δ^N has a feasibility with a slightly relaxed bound, by a factor of $\sqrt{w_i}$ (see*

Section 4.2 for the general Galerkin optimal control computational strategies, particularly Equations (4.10) and (4.13)).

Remark 2.9. *With the Galerkin formulation outlined here, the initial conditions may be enforced in a weak sense. In other words, ICs may be imposed only up to the order of accuracy of the numerical approximation itself. Consider again Equation (2.59). Integration by parts on the first term results in the Galerkin weak form,*

$$-\int_{-1}^{1} \dot{\phi}_i^N x^N d\xi + \left[\phi_i^N x^N\right]_{-1}^{1} - \frac{\Delta t}{2}\int_{-1}^{1} \phi_i^N f(x^N, u^N) d\xi = 0.$$

In terms of the approximating polynomials (and introducing the true initial condition, $x^N(-1) \to x(-1)$) we have

$$-\sum_{j=0}^{N} \int_{-1}^{1} \dot{\phi}_i^N \phi_j^N d\xi\, \bar{x}^{Nj} - \phi_i^N(-1)x(-1) + \phi_i^N(1)x^N(1) - \frac{\Delta t}{2}\int_{-1}^{1} \phi_i^N f(x^N, u^N) d\xi = 0,$$

for $i = 0, 1, \ldots, N$. Integration by parts, yet again, results in the Galerkin strong form,

$$\sum_{j=0}^{N} D_{ij}\bar{x}^{Nj} + \phi_i^N(-1)\left(\sum_{j=0}^{N} \phi_j^N(-1)\bar{x}^{Nj} - x(-1)\right) \\ - \phi_i^N(1)\left(\sum_{j=0}^{N} \phi_j^N(1)\bar{x}^{Nj} - x^N(1)\right) - c_i = 0. \tag{2.63}$$

Equation (2.63) may be formulated for weak enforcement of ICs by letting $x(-1) = x_0$ and $x^N(1) = x^{NN}$. Additionally, when $Q = N$ LGL quadrature points are used to calculate the Galerkin differentiation matrix and approximate the RHS vector, the system may be simplified as

$$\sum_{j=0}^{N} D_{ij}\bar{x}^{Nj} + \kappa_i - \bar{c}^{Ni} = 0, \tag{2.64}$$

for each $i = 0, 1, \dots, N$, where

$$\kappa_i = \begin{cases} \bar{x}^{N0} - x_0, & i = 0, \\ 0, & i \neq 0. \end{cases}$$

The IC term κ now provides a natural way to introduce initial conditions into the discretization of the dynamics. Again, in order to guarantee feasibility of the discretized problem, the following inequality constraint is suggested,

$$\left\| \sum_{j=0}^{N} D_{ij} \bar{x}^{Nj} + \kappa_i - \bar{c}^{Ni} \right\|_{\infty} \leq \delta^N, \quad i = 0, 1, \dots, N, \tag{2.65}$$

where δ^N is the feasibility tolerance that is dependent on N and the smoothness of x and u (see Section 5.1). Finally, the endpoint conditions may be approximated similarly by

$$\left\| e(\bar{x}^{NN}) \right\|_{\infty} \leq \delta^N.$$

Remark 2.10. *In* [11]*, the equation resulting from dividing Equation (2.64) by w_i is introduced for primal-only closure conditions. However, for feasibility the inequality version of this expression, Equation (2.65), must be used for computational purposes. It should be noted that if the equation in* [11] *is multiplied by w_i first, then relaxed as an inequality bounded by δ^N, the resulting inequality would be in agreement with the feasibility of the Galerkin weak boundary formulation discussed in Section 5.1.2 (see Equation (5.6)).*

Petrov-Galerkin In the *Petrov-Galerkin method* the test and basis functions are different. As with the Bubnov-Galerkin method, these functions can be *modal* or *nodal* in nature. In this section the focus will be on selecting a modal test function and a nodal basis. This will create the framework that will be used in Chapter 4. For this formulation, the selected test functions will be the Legendre polynomials, $\{L_j\}_{j=0}^{N}$, and the Lagrange polynomials

(2.32), $\{\phi_j^N\}_{j=0}^N$, of order N, will be the basis. Again, for this discussion, the LGL node structure will be used for the problem discretization.

The approximation of x and u in the computational space, $\xi \in [-1, 1]$, are given by

$$x^N(\xi) = \sum_{j=0}^{N} \phi_j^N(\xi)\bar{x}^{Nj},$$
$$u^N(\xi) = \sum_{j=0}^{N} \psi_j^N(\xi)\bar{u}^{Nj},$$

where $\{\psi_j^N\}_{j=0}^N$ is any continuous function (not necessarily a polynomial) with the property $\psi_j(\xi_i) = \delta_{ij}$. The expansion coefficients are, $\hat{x}_j = \bar{x}^{Nj}$ and $\hat{u}_j = \bar{u}^{Nj}$, therefore $x^N(\xi_j) = \bar{x}^{Nj}$ and $u^N(\xi_j) = \bar{u}^{Nj}$, for $j = 0, 1, \ldots, N$. Equation (2.58) becomes [44]

$$\int_{-1}^{1} L_i(\xi)\dot{\phi}_j^N(\xi)d\xi \, \bar{x}^{Nj} - \frac{\Delta t}{2}\int_{-1}^{1} L_i(\xi) f(x^N(\xi), u^N(\xi))d\xi = 0, \quad i = 0, 1, \ldots, N,$$

or using matrix-vector notation,

$$\sum_{j=0}^{N} D_{ij}^L \bar{x}^{Nj} - c_i^L = 0, \quad i = 0, 1, \ldots, N.$$

The Galerkin differentiation matrix, D^L, and RHS vector, c^L are defined as

$$D_{ij}^L = \int_{-1}^{1} L_i(\xi)\dot{\phi}_j^N(\xi)d\xi,$$
$$c_i^L = \frac{\Delta t}{2}\int_{-1}^{1} L_i(\xi) f(x^N(\xi), u^N(\xi_k))d\xi,$$

for $i, j = 0, 1, \ldots, N$.

Using LGL quadrature, D^L can be calculated with the relationship

$$D_{ij}^L = \sum_{k=0}^{Q} L_i(\xi_k)\dot{\phi}_j^N(\xi_k)w_k, \quad i, j = 0, 1, \ldots, N,$$

where $\{w_k\}_{k=0}^N$ are the LGL weights given by Equation (2.49). Again, since LGL quadrature rule is exact for polynomial integrands of degree less than or equal to $2N-1$, the numerical integration is done exactly when $Q = N$ LGL integration points are used.

Using LGL quadrature rule, c^L can be approximated by the relationship

$$c_i^L \approx \frac{\Delta t}{2}\sum_{k=0}^{Q} L_i(\xi_k) f(x^N(\xi_k), u^N(\xi_k)) w_k, \quad i = 0, 1, \ldots, N.$$

When $Q = N$ LGL quadrature points are used, the RHS vector approximation, $\bar{c}_L^N$, can be expressed in the simplified form

$$\bar{c}_L^{Ni} = \frac{\Delta t}{2}\sum_{k=0}^{N} L_i(\xi_k) f(\bar{x}^{Nk}, \bar{u}^{Nk}) w_k, \quad i = 0, 1, \ldots, N.$$

Remark 2.11. *Again, if $Q = (N+1)$ integration points are used, the RHS vector will integrate exactly when $f(x(t), u(t))$ is linear in x and u. If f is a nonlinear function, accuracy of integration may be improved by increasing the number of quadrature points Q.*

When $Q = N$ LGL quadrature points are used to calculate D^L and $\bar{c}_L^N$, the system may be simplified as

$$\sum_{j=0}^{N} D_{ij}^L \bar{x}^{Nj} - \bar{c}_L^{Ni} = 0, \quad i = 0, 1, \ldots, N. \tag{2.66}$$

A feasible solution to the equality dynamical constraint may not exist. In order to guarantee feasibility of the discretized problem, the following inequality constraint is suggested,

$$\left\| \sum_{j=0}^{N} D_{ij}^L \bar{x}^{Nj} - \bar{c}_L^{Ni} \right\|_\infty \leq \delta^N, \quad i = 0, 1, \ldots, N,$$

where δ^N is the feasibility tolerance that is dependent on N and the smoothness of x and u (see Chapter 6). The initial conditions and endpoint conditions may be approximated similarly by

$$\left\|\bar{x}^{N0} - x_0\right\|_\infty \leq \delta^N \quad \text{and} \quad \left\|e(\bar{x}^{NN})\right\|_\infty \leq \delta^N.$$

It is clear that in the Petrov-Galerkin formulation (2.66), the differentiation matrix, D^L, and RHS vector, $\bar{c}_L^N$, do not simplify as cleanly as given for the Bubnov-Galerkin formulation (2.60). This inevitably will have negative effects on computational efficiencies. However, casting the problem in the Petrov-Galerkin numerical form will have nice consequences when applied to Galerkin optimal control, as will be shown in Chapter 6.

2.3.2. Spectral Element Methods

Spectral element methods are local (elemental) applications of spectral methods. They combine the flexibility of finite elements with the accuracies associated with spectral methods. This element-based numerical approach is advantageous due to its ability to handle complicated geometries and can be easily formulated for adaptive strategies [16, 59]. In this section, the focus will be on two Galerkin formulations, continuous Galerkin and discontinuous Galerkin element-based methods. Continuous Galerkin techniques were first applied to ordinary differential equations (ODEs) in 1972 by Hulme [14, 15] and a study of global error control was done by Estep et al. [60] in 1994. The first analysis of discontinuous Galerkin methods applied to ODEs was done in 1974 by Reed et al. [61] and an adaptive error control technique was used by Bottcher et al. [62] in 1997. More recently, multi-adaptive continuous Galerkin and discontinuous Galerkin techniques have been studied by Logg and presented in a series of papers [63–65]. The general structure of these element-based Galerkin methods—as well as the mathematical notation used in this dissertation—is provided by Giraldo [44] and discussed in the following sections.

2.3.2.1. Continuous Galerkin

Consider a continuous element-based Galerkin approach to discretizing (2.52). Again, for this discussion, the LGL node structure will be used for the problem discretization. In this approximation, the weak integral form of (2.52) inside each element, Ω_e, takes the form [44]

$$\int_{\Omega_e} \phi_i^{(e)N}(t) \left(\dot{x}^{(e)N}(t) - f(x^{(e)N}(t), u^{(e)N}(t))\right) dt = 0, \tag{2.67}$$

for $e = 1, 2, \ldots, N_e$ and $i = 0, 1, \ldots, N$, where $\Omega = \bigcup_{e=1}^{N_e} \Omega_e$ defines the total domain. The state trajectory, $x(t)$, is approximated inside each element, Ω_e, by interpolating N-th order Lagrange polynomials, $\{\phi_j^{(e)N}(t)\}_{j=0}^N$, at the nodes $\{t_j^{(e)}\}_{j=0}^N$ by the relationship

$$x^{(e)N}(t) = \sum_{j=0}^{N} \phi_j^{(e)N}(t) \bar{x}^{(e)Nj},$$

for $e = 1, 2, \ldots, N_e$, where $\{t_j^{(e)}\}_{j=0}^N$ are the LGL nodes, $\{\xi_j\}_{j=0}^N$, mapped back to the physical space inside each element, Ω_e. Also, let $u^N(t)$ be an interpolating function of $\{\bar{u}^{Nj}\}_{j=0}^N$,

$$u^{(e)N}(t) = \sum_{j=0}^{N} \psi_j^{(e)N}(t) \bar{u}^{(e)Nj},$$

where $\{\psi_j^{(e)N}(t)\}_{j=0}^N$ are any set of continuous functions (not necessarily polynomials) with the property $\psi_j^{(e)N}(t_i) = \delta_{ij}$. Therefore $\bar{x}^{(e)Nj} = x^{(e)N}(t_j^{(e)})$, for $e = 1, 2, \ldots, N_e$ and $j = 0, 1, \ldots, N$, and similarly, $\bar{u}^{(e)Nj} = u^{(e)N}(t_j^{(e)})$. The relationship between the physical time domain, $t \in [t_0, t_f] = \left[t_0^{(1)}, t_N^{(N_e)}\right]$, and the computational space, $\xi \in [-1, 1]$, is given by [44]

$$\xi = \frac{2}{\Delta t^{(e)}} \left(t - t_0^{(e)}\right) - 1 \quad \text{and} \quad d\xi = \frac{2}{\Delta t^{(e)}} dt,$$

and conversely,

$$t = \frac{\Delta t^{(e)}}{2} (\xi + 1) + t_0^{(e)} \quad \text{and} \quad dt = \frac{\Delta t^{(e)}}{2} d\xi,$$

where $\Delta t^{(e)} = t_N^{(e)} - t_0^{(e)}$ is the size of each element, Ω_e, which can be nonuniform in length. The Lagrange polynomial defined on the LGL computational domain is given by

$$\phi_i^N(\xi) = \prod_{\substack{j=0 \\ j\neq i}}^{N} \frac{(\xi - \xi_j)}{(\xi_i - \xi_j)}, \quad i = 0, \dots, N.$$

The state trajectory, x, can now be approximated inside each element, Ω_e, by

$$x^{(e)N}(\xi) = \sum_{j=0}^{N} \phi_j^N(\xi) \bar{x}^{(e)Nj},$$

where $\{\phi_j^N(\xi)\}_{j=0}^N$ are the Lagrange polynomials defined on the LGL grid. Likewise, $u^N(\xi)$ is given by

$$u^{(e)N}(\xi) = \sum_{j=0}^{N} \psi_j^N(\xi) \bar{u}^{(e)Nj},$$

where $\psi_j^N(\xi_i) = \delta_{ij}$.

Remark 2.12. *In this formulation $\bar{x}^{(e)NN} = \bar{x}^{(e+1)N0}$ and $\bar{u}^{(e)NN} = \bar{u}^{(e+1)N0}$, for $e = 1, 2, \dots, N_e - 1$. This continuity condition is a consequence of the global formulation of the problem discussed in Remark 2.13.*

In the computational domain, ξ, the system becomes

$$\int_{-1}^{1} \phi_i^N(\xi) \dot{x}^{(e)N}(\xi) d\xi - \frac{\Delta t^{(e)}}{2} \int_{-1}^{1} \phi_i^N(\xi) f(x^{(e)N}(\xi), u^{(e)N}(\xi)) d\xi = 0,$$

for $e = 1, 2, \dots, N_e$ and $i = 0, 1, \dots, N$, and in terms of the approximating polynomials becomes

$$\sum_{j=0}^{N} \int_{-1}^{1} \phi_i^N(\xi) \dot{\phi}_j^N(\xi) d\xi \, \bar{x}^{(e)Nj} - \frac{\Delta t^{(e)}}{2} \int_{-1}^{1} \phi_i^N f(x^{(e)N}(\xi), u^{(e)N}(\xi)) d\xi = 0.$$

In matrix-vector notation, our system can be expressed as

$$\sum_{j=0}^{N} D_{ij}^{(e)} \bar{x}^{(e)Nj} - c_i^{(e)} = 0,$$

for $e = 1, 2, \ldots, N_e$ and $i = 0, 1, \ldots, N$. The local element $(N+1) \times (N+1)$ Galerkin differentiation matrix, $D^{(e)}$, is defined as

$$D_{ij}^{(e)} = \int_{-1}^{1} \phi_i^N(\xi) \dot{\phi}_j^N(\xi) d\xi, \quad i, j = 0, 1, \ldots, N. \tag{2.68}$$

Using LGL quadrature, $D^{(e)}$, can be calculated with the relationship

$$D_{ij}^{(e)} = \sum_{k=0}^{Q} \phi_i^N(\xi_k) \dot{\phi}_j^N(\xi_k) w_k = \dot{\phi}_j^N(\xi_i) w_i = A_{ij} w_i, \quad i, j = 0, 1, \ldots, N, \tag{2.69}$$

where $\{w_i\}_{i=0}^{N}$ are the LGL weights given by Equation (2.49) and A is the Legendre PS differentiation matrix (2.57). Since LGL quadrature rule is exact for polynomial integrands of degree less than or equal to $2N - 1$, the numerical integration is done exactly when $Q = N$ LGL integration points are used. If $Q = N$ LGL quadrature nodes are used, the approximation to the $(N+1) \times 1$ RHS vector simplifies to

$$c_i^{(e)} \approx \bar{c}^{(e)Ni} = \frac{\Delta t^{(e)}}{2} f(\bar{x}^{(e)Ni}, \bar{u}^{(e)Ni}) w_i,$$

for $e = 1, 2, \ldots, N_e$ and $i = 0, 1, \ldots, N$.

Remark 2.13. *So far, the required objects have been identified to solve the system numerically with element-based Galerkin. However, since nodal basis functions are continuous across element boundaries and LGL nodes include both endpoints, a global solution to our problem can be found. To do this, a global assembly or direct stiffness summation can be done, where the direct stiffness summation operator is* $\bigwedge_{e=1}^{N_e}$. [44]

The global equations to the problem become

$$\sum_{J=1}^{N_p} D_{IJ}\bar{x}^{N_pJ} - \bar{c}^{N_pI} = 0, \quad I = 1, \ldots, N_p. \tag{2.70}$$

The global $N_p \times N_p$ Galerkin differentiation matrix, D_{IJ}, and RHS vector, $\bar{c}^{N_pI}$, are then defined by

$$D_{IJ} = \bigwedge_{e=1}^{N_e} D_{ij}^{(e)} \quad \text{and} \quad \bar{c}^{N_pI} = \bigwedge_{e=1}^{N_e} \bar{c}^{(e)Ni},$$

where $N_p = (N_eN + 1)$ is the total number of grid points. Note that the direct stiffness summation operator, $\bigwedge_{e=1}^{N_e}$, does the mapping $(i, e), (j, e) \rightarrow I, J$ [44]. So for the local differentiation matrix and RHS vector

$$D^{(e)} = \begin{pmatrix} d_{00}^{(e)} & d_{01}^{(e)} & \cdots & d_{0N}^{(e)} \\ d_{10}^{(e)} & d_{11}^{(e)} & \cdots & d_{1N}^{(e)} \\ \vdots & \vdots & \ddots & \vdots \\ d_{N0}^{(e)} & d_{N1}^{(e)} & \cdots & d_{NN}^{(e)} \end{pmatrix} \text{ and } \bar{c}^{(e)N} = \begin{pmatrix} \bar{c}^{(e)N0} \\ \bar{c}^{(e)N1} \\ \vdots \\ \bar{c}^{(e)NN} \end{pmatrix},$$

the direct stiffness summation operations $D_{IJ} = \bigwedge\limits_{e=1}^{2} D_{ij}^{(e)}$ and $\bar{c}^{N_pI} = \bigwedge\limits_{e=1}^{2} \bar{c}^{(e)Ni}$ result in the global $N_p \times N_p$ differentiation matrix and $N_p \times 1$ RHS vector,

$$D = \begin{pmatrix} d_{00}^{(1)} & d_{01}^{(1)} & \cdots & d_{0N}^{(1)} & 0 & \cdots & 0 \\ d_{10}^{(1)} & d_{11}^{(1)} & \cdots & d_{1N}^{(1)} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ d_{N0}^{(1)} & d_{N1}^{(1)} & \cdots & d_{NN}^{(1)} + d_{00}^{(2)} & d_{01}^{(2)} & \cdots & d_{0N}^{(2)} \\ 0 & \cdots & 0 & d_{10}^{(2)} & d_{11}^{(2)} & \cdots & d_{1N}^{(2)} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & d_{N0}^{(2)} & d_{N1}^{(2)} & \cdots & d_{NN}^{(2)} \end{pmatrix} \text{ and } \bar{c}^{N_p} = \begin{pmatrix} \bar{c}^{(1)N0} \\ \bar{c}^{(1)N1} \\ \vdots \\ \bar{c}^{(1)NN} + \bar{c}^{(2)N0} \\ \bar{c}^{(2)N1} \\ \vdots \\ \bar{c}^{(2)NN} \end{pmatrix}.$$

In order to guarantee feasibility of the discretized problem, the following inequality constraint is suggested,

$$\left\| \sum_{J=0}^{N_p} D_{IJ} \bar{x}^{N_p J} - \bar{c}^{N_p I} \right\|_\infty \leq \delta^{N_p}, \quad I = 0, 1, \ldots, N_p,$$

where δ^{N_p} is the feasibility tolerance that is dependent on N_p and the smoothness of x and u (see Section 5.2). The initial conditions and endpoint conditions may be approximated similarly by

$$\left\| \bar{x}^{N_p 0} - x_0 \right\|_\infty \leq \delta^{N_p} \quad \text{and} \quad \left\| e(\bar{x}^{N_p N_p}) \right\|_\infty \leq \delta^{N_p}.$$

Note that although the discretization for Problem (2.67) is element-based, the continuous Galerkin formulation (2.70) is global in nature. This, however, is not the case for the discontinuous element-based Galerkin approach.

2.3.2.2. Discontinuous Galerkin

Consider a discontinuous element-based Galerkin approach to discretizing (2.52). Again, for this discussion, the LGL node structure will be used for the problem discretization. In this approximation, the weak integral form of (2.52) inside each element, Ω_e, takes the form [44]

$$\int_{\Omega_e} \phi_i^{(e)N}(t) \left(\dot{x}^{(e)N}(t) - f(x^{(e)N}(t), u^{(e)N}(t)) \right) dt = 0,$$

for $e = 1, 2, \ldots, N_e$ and $i = 0, 1, \ldots, N$, where $\Omega = \bigcup_{e=1}^{N_e} \Omega_e$ defines the total domain. The state trajectory, $x(t)$, is approximated inside each element, Ω_e, by interpolating N-th order Lagrange polynomials, $\{\phi_j^{(e)N}(t)\}_{j=0}^{N}$, at the nodes $\{t_j^{(e)}\}_{j=0}^{N}$ by the relationship

$$x^{(e)N}(t) = \sum_{j=0}^{N} \phi_j^{(e)N}(t) \bar{x}^{(e)Nj},$$

for $e = 1, 2, \ldots, N_e$, where $\{t_j^{(e)}\}_{j=0}^N$ are the LGL nodes, $\{\xi_j\}_{j=0}^N$, mapped back to the physical space inside each element, Ω_e. Also, let $u^N(t)$ be an interpolating function of $\{\bar{u}^{Nj}\}_{j=0}^N$,

$$u^{(e)N}(t) = \sum_{j=0}^{N} \psi_j^{(e)N}(t) \bar{u}^{(e)Nj},$$

where $\{\psi_j^{(e)N}(t)\}_{j=0}^N$ are any set of continuous functions (not necessarily polynomials) with the property $\psi_j^{(e)N}(t_i) = \delta_{ij}$. Therefore $\bar{x}^{(e)Nj} = x^{(e)N}(t_j^{(e)})$, for $e = 1, 2, \ldots, N_e$ and $j = 0, 1, \ldots, N$, and similarly, $\bar{u}^{(e)Nj} = u^{(e)N}(t_j^{(e)})$. The relationship between the physical time domain, $t \in [t_0, t_f] = \left[t_0^{(1)}, t_N^{(N_e)}\right]$, and the computational space, $\xi \in [-1, 1]$, is given by [44]

$$\xi = \frac{2}{\Delta t^{(e)}} \left(t - t_0^{(e)}\right) - 1 \quad \text{and} \quad d\xi = \frac{2}{\Delta t^{(e)}} dt,$$

and conversely,

$$t = \frac{\Delta t^{(e)}}{2} (\xi + 1) + t_0^{(e)} \quad \text{and} \quad dt = \frac{\Delta t^{(e)}}{2} d\xi,$$

where $\Delta t^{(e)} = t_N^{(e)} - t_0^{(e)}$ is the size of each element, Ω_e, which can be nonuniform in length. The Lagrange polynomial defined on the LGL computational domain is given by

$$\phi_i^N(\xi) = \prod_{\substack{j=0 \\ j \neq i}}^{N} \frac{(\xi - \xi_j)}{(\xi_i - \xi_j)}, \quad i = 0, \ldots, N.$$

The state trajectory, x, can now be approximated inside each element, Ω_e, by

$$x^{(e)N}(\xi) = \sum_{j=0}^{N} \phi_j^N(\xi) \bar{x}^{(e)Nj},$$

where $\{\phi_j^N(\xi)\}_{j=0}^N$ are the Lagrange polynomials defined on the LGL grid. Likewise, $u^N(\xi)$ is given by

$$u^{(e)N}(\xi) = \sum_{j=0}^{N} \psi_j^N(\xi)\bar{u}^{(e)Nj},$$

where $\psi_j^N(\xi_i) = \delta_{ij}$. In the computational domain, ξ, the system becomes

$$\int_{-1}^{1} \phi_i^N(\xi)\dot{x}^{(e)N}(\xi)d\xi - \frac{\Delta t^{(e)}}{2}\int_{-1}^{1} \phi_i^N(\xi)f(x^{(e)N}(\xi), u^{(e)N}(\xi))d\xi = 0,$$

for $e = 1, 2, \ldots, N_e$ and $i = 0, 1, \ldots, N$. Integration by parts on the first term yields the weak form relationship

$$-\int_{-1}^{1} \dot{\phi}_i^N(\xi)x^{(e)N}(\xi)d\xi + \left[\phi_i^N(\xi)x^{(e)N}(\xi)\right]_{-1}^{1} - \frac{\Delta t^{(e)}}{2}\int_{-1}^{1} \phi_i^N(\xi)f(x^{(e)N}(\xi), u^{(e)N}(\xi))d\xi = 0,$$

for $e = 1, 2, \ldots, N_e$ and $i = 0, 1, \ldots, N$, and in terms of our approximating polynomials we have

$$\begin{aligned} -\sum_{j=0}^{N} \int_{-1}^{1} \dot{\phi}_i^N(\xi)\phi_j^N(\xi)d\xi \; \bar{x}^{(e)Nj} + \sum_{j=0}^{N} \left[\phi_i^N(\xi)\phi_j^N(\xi)\right]_{-1}^{1} \; \bar{x}_j^{(*)} \\ -\frac{\Delta t^{(e)}}{2}\int_{-1}^{1} \phi_i^N(\xi)f(x^{(e)N}(\xi), u^{(e)N}(\xi))d\xi = 0. \end{aligned}$$

Remark 2.14. *With the discontinuous element-based Galerkin approach, we let $\dot{x}$, u and the basis functions be discontinuous across element edges. A numerical flux term $\bar{x}^{(*)}$ acts as a jump condition between elements* [44]. *Here, we consider the centered flux relationship, $\bar{x}^{(*)} = \frac{1}{2}\left(\bar{x}^{(e)} + \bar{x}^{(q)}\right)$, proposed by Delfour et al.* [66], *where e and q denote the element and its neighbor, respectively.*

Integrating by parts, yet again, results in the Galerkin strong form relationship

$$\sum_{j=0}^{N} \int_{-1}^{1} \phi_i^N(\xi)\dot{\phi}_j^N(\xi)d\xi\, \bar{x}^{(e)Nj} + \eta_i^{(e)} - \frac{\Delta t^{(e)}}{2} \int_{-1}^{1} \phi_i^N(\xi) f(x^{(e)N}(\xi), u^{(e)N}(\xi))d\xi = 0,$$

for $e = 1, 2, \ldots, N_e$ and $i = 0, 1, \ldots, N$. Since LGL nodes are used, the boundary term, $\eta^{(e)}$, may be simplified as

$$\eta_i^{(1)} = \begin{cases} \frac{1}{2}\left(\bar{x}^{(2)N0} - \bar{x}^{(1)NN}\right), & i = N, \\ 0, & i \neq N, \end{cases}$$

$$\eta_i^{(N_e)} = \begin{cases} \frac{1}{2}\left(\bar{x}^{(N_e)N0} - \bar{x}^{(N_e-1)NN}\right), & i = 0, \\ 0, & i \neq 0, \end{cases}$$

for elements $\Omega_e = \Omega_1$ and Ω_{N_e}, respectively, and for each other element $(\Omega_e \neq \Omega_1, \Omega_{N_e})$ we have

$$\eta_i^{(e)} = \begin{cases} \frac{1}{2}\left(\bar{x}^{(e)N0} - \bar{x}^{(e-1)NN}\right), & i = 0, \\ \frac{1}{2}\left(\bar{x}^{(e+1)N0} - \bar{x}^{(e)NN}\right), & i = N, \\ 0, & i \neq 0, N. \end{cases}$$

In matrix-vector notation, our system may be expressed as

$$\sum_{j=0}^{N} D_{ij}^{(e)} \bar{x}_j^{(e)} + \eta_i^{(e)} - c_i^{(e)} = 0,$$

for $e = 1, 2, \ldots, N_e$ and $i = 0, 1, \ldots, N$, where the local element $(N+1) \times (N+1)$ Galerkin differentiation matrix, $D^{(e)}$, is the same as that defined in (2.68) and (2.69). If $Q = N$ LGL quadrature nodes are used, the approximation to the $(N+1) \times 1$ RHS vector simplifies to

$$c_i^{(e)} \approx \bar{c}^{(e)Ni} = \frac{\Delta t^{(e)}}{2} f(\bar{x}^{(e)Ni}, \bar{u}^{(e)Ni}) w_i,$$

for $e = 1, 2, \ldots, N_e$ and $i = 0, 1, \ldots, N$. The local discontinuous formulation therefore becomes

$$\sum_{j=0}^{N} D_{ij}^{(e)} \bar{x}_j^{(e)} + \eta_i^{(e)} - \bar{c}^{(e)Ni} = 0, \tag{2.71}$$

for $e = 1, 2, \ldots, N_e$ and $i = 0, 1, \ldots, N$. In order to guarantee feasibility of the discretized problem, the following inequality constraint is suggested,

$$\left\| \sum_{j=0}^{N} D_{ij}^{(e)} \bar{x}^{(e)Nj} + \eta_i^{(e)} - \bar{c}^{(e)Ni} \right\|_\infty \leq \delta^N,$$

for $e = 1, 2, \ldots, N_e$ and $i = 0, 1, \ldots, N$, where δ^N is the feasibility tolerance that is dependent on N and the smoothness of x and u (see Section 5.3). The initial conditions and endpoint conditions may be approximated similarly by

$$\left\| \bar{x}^{(1)N0} - x_0 \right\|_\infty \leq \delta^N \quad \text{and} \quad \left\| e(\bar{x}^{(N_e)NN}) \right\|_\infty \leq \delta^N.$$

Note that unlike the continuous element-based Galerkin approach that can easily be formulated for a global solution (2.70), the discontinuous Galerkin formulation (2.71) is purely local in nature. The communication between elements is done only by the boundary term, $\eta^{(e)}$. It is therefore easy to see that the discontinuous Galerkin formulation is easy to parallelize for computational efficiency. Additionally, the flexibility and discontinuous nature of the formulation lends itself to problems with complex geometries and discontinuous solutions.

Global spectral method techniques, specifically, collocation (or PS methods) have become the method of choice for discretizing system dynamics in a number of direct methods for optimal control, such as the Legendre PS method. Element-based collocation techniques have also been investigated for the use in direct methods for optimal control by Ross et al. [12]. These methods have gained attention due to their flexibility as well as computational efficiency. In Chapter 5, we will further investigate the use of the element-based Galerkin formulations for optimal control. However, we will first consider additional motivation for the use of the weak integral formulation in Chapter 3. This will lead to the creation of a family of Galerkin-based formulations called, Galerkin optimal control.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 3: MOTIVATION FOR GALERKIN OPTIMAL CONTROL

PS methods for optimal control have been shown to be good all-round methods for solving nonlinear optimal control problems. In particular, the Legendre PS method has gained much attention in recent years. As mentioned previously, two highlights are the successful use of Legendre PS method for the first ever zero-propellant attitude maneuver of the International Space Station [4] and the first ever minimum-time rotational maneuver performed in orbit by a NASA space telescope called TRACE [8]. The formulation of the Legendre PS method is provided in Section 3.2, but first consider the following problem of optimal control.

3.1. Problem B (Bolza Problem)

Determine the state-control function pair, $t \mapsto (x, u) \in \mathbb{R}^{N_x} \times \mathbb{R}^{N_u}$, that minimizes the cost functional

$$J = \int_{-1}^{1} F(x(t), u(t))dt + E(x(-1), x(1)), \tag{3.1}$$

subject to the dynamics

$$\dot{x}(t) = f(x(t), u(t)), \tag{3.2}$$

endpoint conditions

$$e(x(-1), x(1)) = 0, \tag{3.3}$$

and mixed state-control path conditions

$$h(x(t), u(t)) \leq 0. \tag{3.4}$$

It is assumed that $F : \mathbb{R}^{N_x} \times \mathbb{R}^{N_u} \to \mathbb{R}$, $E : \mathbb{R}^{N_x} \times \mathbb{R}^{N_x} \to \mathbb{R}$, $f : \mathbb{R}^{N_x} \times \mathbb{R}^{N_u} \to \mathbb{R}^{N_x}$, $e : \mathbb{R}^{N_x} \times \mathbb{R}^{N_x} \to \mathbb{R}^{N_e}$, $h : \mathbb{R}^{N_x} \times \mathbb{R}^{N_u} \to \mathbb{R}^{N_h}$ are Lipschitz continuous with respect to their argument. It is also assumed that an optimal solution $(x^*(\cdot), u^*(\cdot))$ exists. Additional assumptions related to the smoothness of $x^*(\cdot)$ and $u^*(\cdot)$ are provided in the feasibility and consistency theorems included in Chapters 4, 5 and 6.

3.2. Legendre Pseudospectral Method

In the Legendre PS method approximation to Problem B, the states are approximated with globally interpolating N-th order Lagrange polynomials defined on LGL grid, $\{t_j\}_{j=0}^N$. Recall that the LGL points are defined by $t_0 = -1 < t_1 < \cdots < t_N = 1$ and are the roots of Equation (2.42). The state trajectory, $x(t)$, is approximated by

$$x(t) \approx x^N(t) = \sum_{j=0}^{N} \phi_j^N(t) \bar{x}^{Nj}.$$

Let

$$\bar{x}^{Nj} \approx x(t_j), \quad j = 0, 1, \ldots, N,$$

and similarly, $\bar{u}^{Nj} \approx u(t_j)$. The Lagrange polynomials, $\{\phi_j^N\}_{j=0}^N$, of order N, are given by Equation (2.32), and have the property, $\phi_j^N(t_i) = \delta_{ij}$, for $i, j = 0, 1, \ldots, N$.

In the Legendre PS method, a solution to the differential equation $\dot{x} - f(x, u) = 0$ may be approximated at the LGL nodes with the following formulation

$$\sum_{j=0}^{N} A_{ij} \bar{x}^{Nj} - f(\bar{x}^{Ni}, \bar{u}^{Ni}) = 0, \quad i = 0, 1, \ldots, N, \tag{3.5}$$

since, the derivative of $x^N(t)$ at each LGL point $\{t_i\}_{i=0}^N$ is exactly equal to

$$\dot{x}^N(t_i) = \sum_{j=0}^{N} A_{ij}\bar{x}^{Nj},$$

for any polynomial with degree less than or equal to N [46], where A is the Legendre PS differentiation matrix (2.57). A feasible solution to the equality dynamical constraint may not exist. In order to guarantee feasibility of the discretized problem, Gong et al. [5], suggest the following inequality constraint,

$$\left\| \sum_{j=0}^{N} A_{ij}\bar{x}^{Nj} - f(\bar{x}^{Ni}, \bar{u}^{Ni}) \right\|_\infty \leq \delta^N, \quad i = 0, 1, \ldots, N.$$

Remark 3.1. *Note that δ^N is the feasibility tolerance and is dependent on N and the smoothness of x and u. For $x \in W^{m,\infty}$ (see Appendix C), $m \geq 2$ and $u \in C^0[-1,1]$, it has been proven by Gong et al.* [5] *that $\delta^N = (N-1)^{\frac{3}{2}-m}$.*

The endpoint conditions and path constraints are approximated similarly by

$$\begin{aligned} \left\| e(\bar{x}^{N0}, \bar{x}^{NN}) \right\|_\infty &\leq \delta^N, \\ h(\bar{x}^{Ni}, \bar{u}^{Ni}) &\leq \delta^N \cdot \mathbf{1}, \quad i = 0, 1, \ldots, N, \end{aligned}$$

where $\mathbf{1}$ denotes $[1, \ldots, 1]^T$. Lastly, the cost functional $J[x(\cdot), u(\cdot)]$ is approximated by LGL quadrature rule,

$$J[x(\cdot), u(\cdot)] \approx \bar{J}^N(\bar{x}^N, \bar{u}^N) = \sum_{i=0}^{N} F(\bar{x}^{Ni}, \bar{u}^{Ni}) w_i + E(\bar{x}^{N0}, \bar{x}^{NN}),$$

where $\bar{x}^N = \left[\bar{x}^{N0},\ \bar{x}^{N1}, \ldots, \bar{x}^{NN}\right]$, $\bar{u}^N = \left[\bar{u}^{N0},\ \bar{u}^{N1}, \ldots, \bar{u}^{NN}\right]$ and $\{w_i\}_{i=0}^N$ are the LGL weights (2.49) associated with the LGL points, $\{t_i\}_{i=0}^N$. To allow for a practical search area for the optimal solution the following constraints are added

$$\{\bar{x}^{Ni} \in \boldsymbol{X}, \bar{u}^{Ni} \in \boldsymbol{U},\ i = 0, 1, \ldots, N\},$$

where $\boldsymbol{X}$ and $\boldsymbol{U}$ are the search regions that contain the optimal solution of the discretized nonlinear optimization.

The resulting optimization problem can be solved using existing NLP algorithms. A feasible solution can be found that satisfies the tolerances specified in the NLP by adjusting the order of polynomial used in the approximation. The theoretical underpinnings of the Legendre PS method have been studied in great detail over the last two decades. Theorems for feasibility, consistency and convergence of the Legendre PS method approximations can be found in [3, 5, 6, 67, 68]. Although the Legendre PS method has been shown to produce accurate solutions on a wide variety of optimal control problems, it has proven to be a challenging task to modify this method to efficiently solve multi-scale problems, one for which the state(s) and control(s) evolve at different timescales. An example of such a problem is given next.

Example 3.1. Consider the following boundary value problem given by Williams [69] of minimizing the cost function

$$J = \frac{1}{2} \int_0^{t_f} u^2 dt, \tag{3.6}$$

subject to the dynamics

$$\dot{x}_1(t) = x_2 \quad \text{and} \quad \dot{x}_2(t) = C \sin(kt) + u, \tag{3.7}$$

and with boundary conditions

$$x_1(0) = 0, \quad x_2(0) = 0, \quad x_1(t_f) = 1 \quad \text{and} \quad x_2(t_f) = 0. \tag{3.8}$$

The analytic solution to this problem is given by

$$x_1(t) = -\frac{C}{k^2}\sin(kt) - c_1\frac{t^3}{6} - c_2\frac{t^2}{2} + \frac{C}{k}t, \tag{3.9}$$

$$x_2(t) = -\frac{C}{k}\cos(kt) - c_1\frac{t^2}{2} - c_2 t + \frac{C}{k}, \tag{3.10}$$

$$u(t) = -c_1 t - c_2, \tag{3.11}$$

obtained via Pontryagin's maximum principle. The constants are defined as

$$c_1 = -\frac{C}{t_f^2}\left(\frac{C}{k}\cos(kt_f) - \frac{C}{k}\right) + \frac{12}{t_f^3}\left(1 + \frac{C}{k^2}\sin(kt_f) - \frac{C}{k}t_f\right), \tag{3.12}$$

$$c_2 = \frac{2}{t_f}\left(\frac{C}{k}\cos(kt_f) - \frac{C}{k}\right) + \frac{6}{t_f^2}\left(1 + \frac{C}{k^2}\sin(kt_f) - \frac{C}{k}t_f\right), \tag{3.13}$$

$t_f = 10$, $C = 0.1$ and $k = 8$. This problem was solved using the Legendre PS method with optimality and feasibility tolerances of 5×10^{-5} and 5×10^{-4}, respectively. The exact solution was used as an initial guess. Figure 12 shows the Legendre PS method approximations of order, $N = 50$.

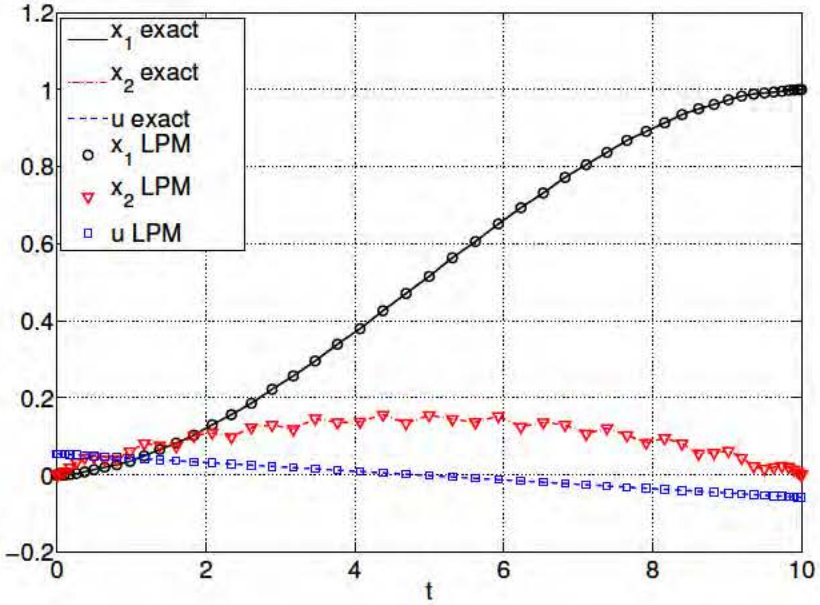
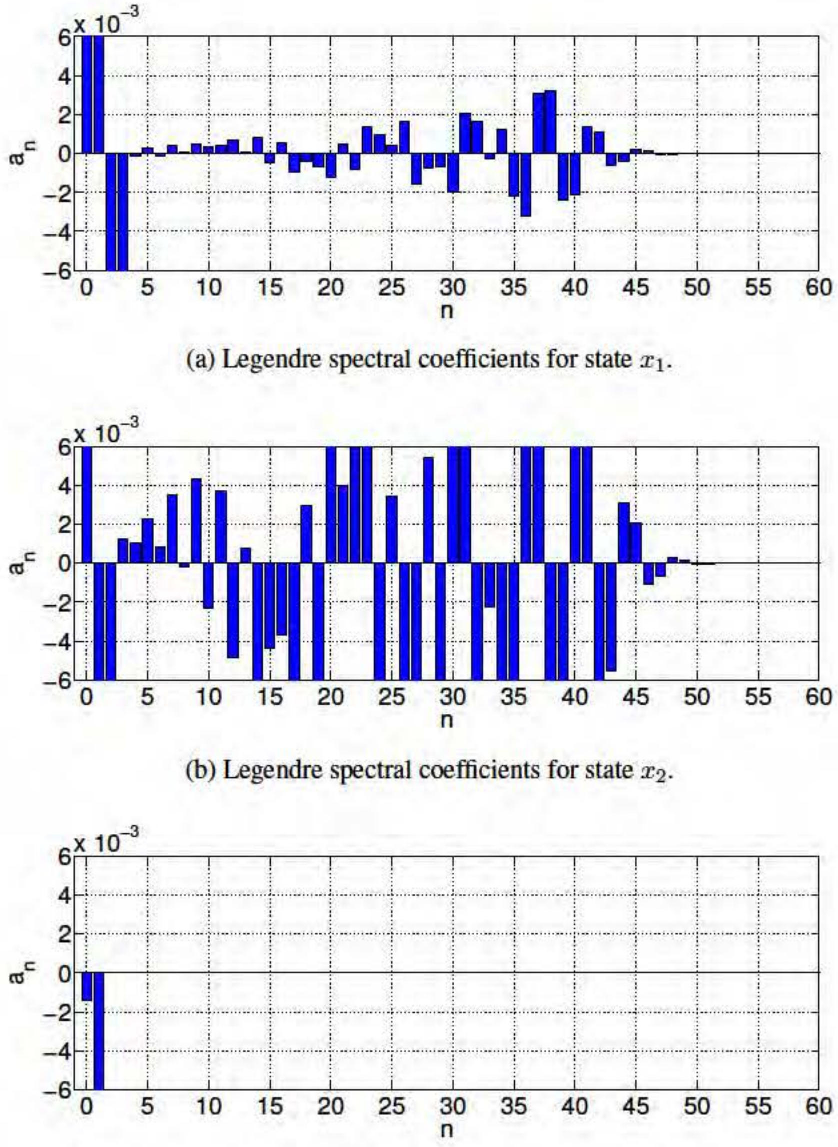


Figure 12: Exact solution and Legendre PS method approximation with $N = 50$ for Example 3.1.

From Figure 12, it is apparent that x_1 and x_2 evolve on different timescales. This is confirmed by viewing the Legendre spectral coefficients of x_1 and x_2 presented in Figure 13. Note the difference in magnitude of the x_1 and x_2 Legendre spectral coefficients, particularly between $n = 5$ and 40.



(a) Legendre spectral coefficients for state x_1.

(b) Legendre spectral coefficients for state x_2.

(c) Legendre spectral coefficients for control u.

Figure 13: Legendre spectral coefficients for x_1, x_2 and u for Example 3.1.

The difference in evolution of the x_1 and x_2 system dynamics suggests that problem (3.6)–(3.8) may be approximated more efficiently using a multi-scale numerical technique, where slow state, x_1, and fast state, x_2 are discretized on different timescales. Consider the following general multi-scale optimal control problem, in which the slow and fast states,

$x_s(t)$ and $x_f(t)$, are associated with the slow and fast dynamics, respectively. This modified Problem B is presented as Problem B̃.

3.3. Problem B̃ (Multi-scale Bolza Problem)

Problem B̃. Determine the state-control function, $t \mapsto (x_s, x_f, u) \in \mathbb{R}^{N_{x_s}} \times \mathbb{R}^{N_{x_f}} \times \mathbb{R}^{N_u}$, that minimizes the cost functional

$$J[x_s(\cdot), x_f(\cdot), u(\cdot)] = \int_{-1}^{1} F(x_s(t), x_f(t), u(t))dt + E(x_s(-1), x_s(1), x_f(-1), x_f(1)),$$

subject to the dynamics,

$$\dot{x}_s(t) = f(x_s(t), x_f(t), u(t)),$$
$$\dot{x}_f(t) = g(x_s(t), x_f(t), u(t)),$$

endpoint conditions,

$$e(x_s(-1), x_s(1), x_f(-1), x_f(1)) = 0,$$

and mixed state-control path conditions,

$$h(x_s(t), x_f(t), u(t)) \leq 0.$$

It is assumed that F: $\mathbb{R}^{N_{x_s}} \times \mathbb{R}^{N_{x_f}} \times \mathbb{R}^{N_u} \to \mathbb{R}$, E: $\mathbb{R}^{N_{x_s}} \times \mathbb{R}^{N_{x_s}} \times \mathbb{R}^{N_{x_f}} \times \mathbb{R}^{N_{x_f}} \to \mathbb{R}$, f: $\mathbb{R}^{N_{x_s}} \times \mathbb{R}^{N_{x_f}} \times \mathbb{R}^{N_u} \to \mathbb{R}$, e: $\mathbb{R}^{N_{x_s}} \times \mathbb{R}^{N_{x_s}} \times \mathbb{R}^{N_{x_f}} \times \mathbb{R}^{N_{x_f}} \to \mathbb{R}^{N_e}$, h: $\mathbb{R}^{N_{x_s}} \times \mathbb{R}^{N_{x_f}} \times \mathbb{R}^{N_u} \to \mathbb{R}^{N_h}$ are Lipschitz continuous with respect to their argument. It is also assumed that an optimal solution $(x_s^*(\cdot), x_f^*(\cdot), u^*(\cdot))$ exists.

A number of methods have been investigated for solving multi-scale problems such as Problem B̃, by casting the slow and fast dynamics of the problem onto different timescales. Recently, Desai et al. [70] and Williams [69] provided varied techniques.

While Desai et al. use an elemental approach where fast and slow dynamics are treated with similar order polynomials within different size subintervals and Williams uses a technique where the slow dynamics are approximated with a weak formulation. Additionally, in [71], Gong et al. investigate the use of a Tau-like method to discretize the slow dynamics, after discounting a straightforward modified Legendre PS method approach. This modified Legendre PS method will be presented next for discussion purposes.

3.4. A Modified Legendre PS Method for Multi-scale Problems

Consider the following modified Legendre PS method approach to solving Problem $\tilde{\text{B}}$. The states and controls are approximated with globally interpolating Lagrange polynomials on different LGL timescales. The slow state, $x_s(t)$, is approximated on sparse grid $\{\tau_j\}_{j=0}^{M}$ while the fast state, $x_f(t)$, on dense grid $\{t_j\}_{j=0}^{N}$, where $M < N$. The slow and fast states are defined by the following approximating polynomials

$$x_s(t) \approx x_s^M(t) = \sum_{j=0}^{M} \phi_j^M(t)\bar{x}_s^{Mj},$$
$$x_f(t) \approx x_f^N(t) = \sum_{j=0}^{N} \phi_j^N(t)\bar{x}_f^{Nj},$$

where the Lagrange polynomials $\{\phi_j^M(t)\}_{j=0}^{M}$ and $\{\phi_j^N(t)\}_{j=0}^{N}$ are defined on grids $\{\tau_j\}_{j=0}^{M}$ and $\{t_j\}_{j=0}^{N}$, respectively. Let

$$\bar{x}_s^{Mj} \approx x_s(\tau_j), \quad j = 0, 1, \ldots, M,$$
$$\bar{x}_f^{Nj} \approx x_f(t_j), \quad j = 0, 1, \ldots, N,$$

and similarly, $\bar{u}^{Nj} \approx u(t_j)$, for $j = 0, 1, \ldots, N$.

Remark 3.2. *For simplicity, the control variable, $u(t)$, is approximated on the dense grid $\{t_j\}_{j=0}^{N}$, however this need not be the case. Modifications may be made to this method to cast the control onto a unique grid, such as sparse grid $\{\tilde{\tau}_j\}_{j=0}^{\tilde{M}}$, where $\tilde{M} < N$* [71].

A solution to the differential equations $\dot{x}_s = f(x_s, x_f, u)$ and $\dot{x}_f = g(x_s, x_f, u)$ may be approximated by discretizing the slow dynamics over the dense grid with the following formulation

$$\begin{aligned}\sum_{j=0}^{M} A_{ij}^{NM} \bar{x}_s^{Mj} - f(\hat{x}_s^{Ni}, \bar{x}_f^{Ni}, \bar{u}^{Ni}) = 0, \quad i = 0, 1, \ldots, N,\\ \sum_{j=0}^{N} A_{ij}^{NN} \bar{x}_f^{Nj} - g(\hat{x}_s^{Ni}, \bar{x}_f^{Ni}, \bar{u}^{Ni}) = 0, \quad i = 0, 1, \ldots, N,\end{aligned}$$

where A^{NN} is the standard $(N+1) \times (N+1)$ Legendre PS differentiation matrix (2.57) and A^{NM} is the $(N+1) \times (M+1)$ Legendre PS differentiation transformation matrix (2.39). The slow state approximation projected to the dense grid, $\hat{x}_s^N$, may be calculated by the linear mapping $T_{ij}^{NM} = \phi_j^M(t_i)$ with the relationship

$$\hat{x}_s^{Ni} = \sum_{j=0}^{n} T_{ij}^{NM} \bar{x}_s^{Mj},$$

for $i = 0, 1, \ldots, N$, where T^{NM} is the $(N+1) \times (M+1)$ transformation matrix (2.37).

Remark 3.3. *Projecting the slow dynamics onto the dense grid provides a way of capturing the high frequency information of the fast state* [71].

The dynamical constraints therefore become

$$\begin{aligned}\left\| \sum_{j=0}^{M} A_{ij}^{NM} \bar{x}_s^{Mj} - f(\hat{x}_s^{Ni}, \bar{x}_f^{Ni}, \bar{u}^{Ni}) \right\|_\infty \leq \delta^N, \quad i = 0, 1, \ldots, N,\\ \left\| \sum_{j=0}^{N} A_{ij}^{NN} \bar{x}_f^{Nj} - g(\hat{x}_s^{Ni}, \bar{x}_f^{Ni}, \bar{u}^{Ni}) \right\|_\infty \leq \delta^N, \quad i = 0, 1, \ldots, N,\end{aligned}$$

where δ^N is the feasibility tolerance. The endpoint conditions and path constraints are approximated similarly by

$$\left\|e(\bar{x}_s^{M0}, \bar{x}_s^{MM}, \bar{x}_f^{N0}, \bar{x}_f^{NN})\right\|_\infty \leq \delta^N,$$

$$h(\hat{x}_s^{Ni}, \bar{x}_f^{Ni}, \bar{u}^{Ni}) \leq \delta^N \cdot \mathbf{1}, \quad i = 0, 1, \ldots, N.$$

Lastly, the cost functional $J[x(\cdot), u(\cdot)]$ is approximated by the LGL quadrature rule,

$$J[x(\cdot), u(\cdot)] \approx \bar{J}^N(\bar{x}_s^M, \bar{x}_f^N, \bar{u}^N) = \sum_{i=0}^{N} F(\hat{x}_s^{Ni}, \bar{x}_f^{Ni}, \bar{u}^{Ni})w_i + E(\bar{x}_s^{M0}, \bar{x}_s^{MM}, \bar{x}_f^{N0}, \bar{x}_f^{NN}),$$

where $\{w_i\}_{i=0}^N$ are the LGL weights (2.49) associated with the LGL points, $\{t_i\}_{i=0}^N$ and

$$\bar{x}_s^M = \left[\bar{x}_s^{M0},\ \bar{x}_s^{M1}, \ldots, \bar{x}_s^{MM}\right], \quad \bar{x}_f^N = \left[\bar{x}_f^{N0},\ \bar{x}_f^{N1}, \ldots, \bar{x}_f^{NN}\right]$$
$$\text{and} \quad \bar{u}^N = \left[\bar{u}^{N0},\ \bar{u}^{N1}, \ldots, \bar{u}^{NN}\right].$$

To allow for a practical search area for the optimal solution the following constraints are included: $\bar{x}_s^M \in \boldsymbol{X}_s$, $\bar{x}_f^N \in \boldsymbol{X}_f$ and $\bar{u}^N \in \boldsymbol{U}$, where $\boldsymbol{X}_s$, $\boldsymbol{X}_f$ and $\boldsymbol{U}$ are the search regions that contain the optimal solution of the discretized nonlinear optimization.

Example 3.1 (continued)**.** Consider again problem (3.6)–(3.8) solved with the proposed multi-scale Legendre PS method. The following analysis follows that given by Gong et al. in [71]. Here we discretize the slow state, x_1, on LGL grid, $\{\tau_j\}_{j=0}^{N_{x_1}}$, and fast state, x_2 and control, u, on LGL grid $\{t_j\}_{j=0}^{N_{x_2}}$, such that $N_{x_1} < N_{x_2}$. This problem was solved with optimality and feasibility tolerances of 5×10^{-4} & 5×10^{-3}, respectively. The exact solution was used as an initial guess. Figure 14 shows the visual accuracy of the multi-scale Legendre PS method approximations with $N_{x_1} = 40$, $N_{x_2} = 50$ and $N_u = 50$. A decrease in the approximation order of the slow state by 10 causes a significant decrease in the accuracy of the overall approximation. This is particularly apparent in the approximation of the control, u, in Figure 14.

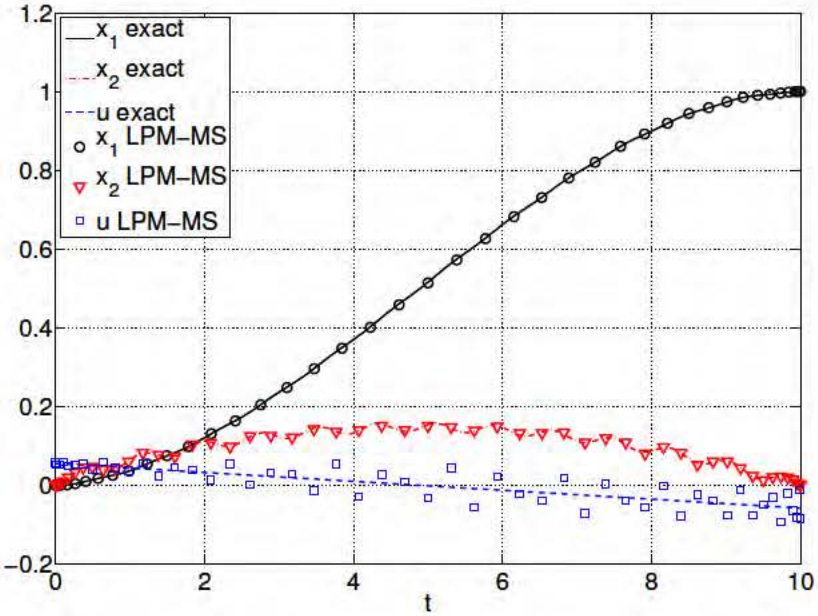


Figure 14: Exact solution and multi-scale Legendre PS method approximation with $N_{x_1} = 40$, $N_{x_2} = 50$ and $N_u = 50$ for Example 3.1.

Also note that if now the control, u, is cast on a unique LGL grid $\{\tilde{\tau}_j\}_{j=0}^{N_u}$, such that $N_u < N_{x_2}$, the NLP becomes infeasible. However, if u is cast on a unique LGL grid such that $N_u \geq N_{x_2}$, an accurate solution is obtained.

Although the Legendre PS method has been shown to produce accurate solutions on a wide variety of optimal control problems, approximating the derivative of a function using a standard PS differentiation matrix may introduce errors into the approximation. This may be an issue when using a multi-scale approach such as the one presented in Section 3.4. It should be mentioned, however, that the Tau-like method of Gong et al. [71] produces accurate solutions for this multi-scale approximation. However, to understand what happened with the straightforward multi-scale approach, we look to Jackson's Theorem.

3.5. Jackson's Theorem

Jackson's Theorem allows for a bounding of the spectral coefficients (2.26), $\{a_n\}_{n=0}^{\infty}$, of a function, $H(t)$, in terms of the function's derivative, $h(t)$.

Lemma 3.1 (Jackson's Theorem)**.** [72] *Let $h(\xi)$ be of bounded variation in $[-1, 1]$. Define*

$$H(\xi) = H(-1) + \int_{-1}^{\xi} h(s)ds,$$

then $\{a_n\}_{n=0}^{\infty}$, the sequence of the spectral coefficients of $H(\xi)$ satisfies the following inequality

$$|a_n| < \frac{6}{\sqrt{\pi}}(U(h(\xi)) + V(h(\xi)))\frac{1}{n^{3/2}},$$

for $n \geq 1$ where $U(h(\xi))$ is the least upper bound of $|h(\xi)|$ and $V(h(\xi))$ is the total variation of $h(\xi)$ (see Appendix A).

To see how this theorem affects a PS approximation of a function's derivative, let H be the approximating polynomial error of a function, let h be its derivative, and let $\{a_n\}_{n=N+1}^{\infty}$ be the spectral coefficients of H. Jackson's Theorem says that even though $\sum\limits_{n=N+1}^{\infty} |a_n|$ may be very small (such as in the tail of the spectral coefficients dropped from an approximation) the error in the approximation of the derivative may be relatively large,

$$|a_n|\frac{\sqrt{\pi}}{6}n^{3/2} < (U(h(\xi)) + V(h(\xi))).$$

This factor of $n^{3/2}$ could potentially add unnecessary errors when approximating a system's dynamics using a standard differentiation matrix.

Remark 3.4. *This idea is further understood by considering the following estimates on the approximation of any function $\zeta(t) \in H^2$ (see Appendix C). Consider $\zeta(t)$ approximated* *by the truncated Legendre series*

$$p^N(t) = \sum_{j=0}^{N} a_j L_j(t),$$

where L_j is the Legendre polynomial of order j and $\{a_j\}_{j=0}^N$ are spectral coefficients of ζ. The error estimate between $\zeta(t)$ and its approximation, $p^N(t)$, is given by

$$\left\|\zeta(t) - p^N(t)\right\|_{L^\infty} = \left\|\sum_{j=N+1}^{\infty} a_j L_j(t)\right\|_{L^\infty} \leq \sum_{j=N+1}^{\infty} |a_j| \|L_j(t)\|_{L^\infty} \leq \sum_{j=N+1}^{\infty} |a_j|,$$

due to the property of the Legendre polynomials [46], $|L_j(t)| \leq 1$, $t \in [-1, 1]$ *(see Appendix A for definition of L^∞-norm). Additionally, the following estimate is provided by* [46]

$$\left\|\zeta(t) - p^N(t)\right\|_{L^\infty} \leq C_1 C_0 N^{-\frac{3}{2}}, \tag{3.14}$$

where C_1 is a constant independent of N and $C_0 = V(\zeta^{(2)})$, the total variation of $\zeta^{(2)}$ (see Appendix A). Now consider the error estimate between $\dot{\zeta}(t)$ and its approximation

$$\dot{p}^N(t) = \sum_{j=0}^{N} a_j \dot{L}_j(t),$$

given by

$$\left\|\dot{\zeta}(t) - \dot{p}^N(t)\right\|_{L^\infty} = \left\|\sum_{j=N+1}^{\infty} a_j \dot{L}_j(t)\right\|_{L^\infty} \leq \sum_{j=N+1}^{\infty} |a_j| \left\|\dot{L}_j(t)\right\|_{L^\infty} \leq \frac{1}{2} \sum_{j=N+1}^{\infty} |a_j| j(j+1),$$

due to the property of the Legendre polynomials [46], $\left|\dot{L}_j(t)\right| \leq \frac{1}{2} j(j+1)$, $t \in [-1, 1]$. *Additionally, the following estimate is provided by* [46]

$$\left\|\dot{\zeta}(t) - \dot{p}^N(t)\right\|_{L^\infty} \leq C_3 C_2 N^{\frac{1}{2}}, \tag{3.15}$$

where C_3 is a constant independent of N and $C_2 = |\zeta|_{H^{2;N}}$, the Sobolev seminorm of ζ (defined in Appendix C). The disparity between estimates (3.14) and (3.15) are clear and thus $p^N(t)$ and $\dot{p}^N(t)$ may converge at different rates. In fact, while estimate (3.14) proves convergence of $p^N(t)$, estimate (3.15) shows that $\dot{p}^N(t)$ may not converge at all. This disproportionate convergence behavior may have compounding effects when using the multi-scale approach for optimal control (introduced in Section 3.4).

Example 3.1 (continued)**.** Consider again problem (3.6)–(3.8) solved with the proposed multi-scale Legendre PS method and $N_{x_1} = 40$, $N_{x_2} = 50$ and $N_u = 50$. If now the optimality and feasibility tolerances are relaxed and decreased to 5×10^{-2} and 5×10^{-1}, respectively, the NLP constraints are satisfied and an accurate approximation of the states and control is obtained. In the context of Jackson's Theorem and Remark 3.4, this should not be a surprise. From Figure 13, the Legendre spectrum of the dropped x_1 modes consist of coefficients with magnitudes of $O(10^{-3})$. In fact with the lower optimality and feasibility tolerances, the multi-scale Legendre PS method can now produce accurate solutions for lower order approximations of x_1, such as $N_{x_1} = 10$ (with N_{x_2} and $N_u = 50$). However, if a reduction in the control approximation is also the goal such as, $N_u \leq 40$ (with $N_{x_1} = 10$ and $N_{x_2} = 50$), a further reduction in the optimality and feasibility tolerances are required in order to satisfy the NLP constraints.

The consequences of Jackson's Theorem on multi-scale PS methods for optimal control are significant. Certainly, the class of problems that can obtain an advantage from this approach is limited. In general we can only hope to benefit from multi-scale PS when reducing the polynomial order of system variables that have extremely small Legendre expansion coefficients at the tail of the spectrum. This will require us to use a different approach if we hope to target a larger class of optimal control problems. Proposition 3.1 highlights the advantage of an alternate method of discretizing the system dynamics, one in which the derivative of higher order terms does not disproportionally add to the overall error of the approximation.

3.6. Motivation for the Weak Integral Formulation

Proposition 3.1. *Let $L_j(t)$ be the Legendre polynomial of order j. Suppose*

$$\epsilon^N(t) = \sum_{j=N+1}^{\infty} a_j L_j(t), \tag{3.16}$$

$$\dot{\epsilon}^N(t) = \sum_{j=N+1}^{\infty} a_j \dot{L}_j(t), \tag{3.17}$$

are both uniformly convergent on $[-1, 1]$, then

$$\int_{-1}^{1} L_i(t)\dot{\epsilon}^N(t)\mathrm{d}t = 2 \sum_{\substack{j=N+1 \\ i+j \text{ odd}}}^{\infty} a_j, \tag{3.18}$$

for all $0 \le i \le N$.

Proof.

$$\begin{aligned}\int_{-1}^{1} L_i(t)\dot{\epsilon}^N(t)\mathrm{d}t &= \sum_{j=N+1}^{\infty} a_j \int_{-1}^{1} L_i(t)\dot{L}_j(t)dt \\ &= \sum_{j=N+1}^{\infty} a_j \left(L_i(t)L_j(t) \mid_{-1}^{1} - \int_{-1}^{1} \dot{L}_i(t)L_j(t)dt \right)\end{aligned}$$

Since the order of each $\{\dot{L}_i(t)\}_{i=0}^{N}$ is less than N and the order of each $\{L_j(t)\}_{N+1}^{\infty}$ is bigger than N, the orthogonality of the Legendre polynomials implies

$$\int_{-1}^{1} \dot{L}_i(t)L_j(t)dt = 0.$$

Due to the following properties of the Legendre polynomial,

$$L_k(1) = 1 \quad \text{and} \quad L_k(-1) = (-1)^k,$$

Equation (3.18) follows. □

Remark 3.5. *A similar result is found for the case that Lagrange interpolation polynomials are used as test functions. Let $\{\phi_i^N(t)\}_{i=0}^N$ be the Lagrange polynomials of order N, defined on grid $t_0 = -1 < t_1, \ldots, t_{N-1} < t_N = 1$. Also, let $L_j(t)$ be the Legendre polynomial of order j; let $\eta(t)$ and $\dot{\eta}(t)$ be defined by (3.16) and (3.17), respectively. Then*

$$\begin{aligned}\int_{-1}^{1} \phi_i^N(t)\dot{\eta}(t)dt &= \sum_{j=N+1}^{\infty} a_j \int_{-1}^{1} \phi_i^N(t)\dot{L}_j(t)dt \\ &= \sum_{j=N+1}^{\infty} a_j \left(\phi_i^N(t)L_j(t) \mid_{-1}^{1} - \int_{-1}^{1} \dot{\phi}_i^N(t)L_j(t)dt \right).\end{aligned}$$

Since the order of the polynomials $\{\dot{\phi}_i^N(t)\}_{i=0}^N$ is $N-1$ and the order of each $\{L_j(t)\}_{j=N+1}^{\infty}$ is bigger than N, the orthogonality of the Legendre polynomials implies

$$\int_{-1}^{1} \dot{\phi}_i^N(t)L_j(t)dt = 0.$$

Due to the following properties of the Legendre and Lagrange polynomials,

$$\phi_i^N(t_k) = \delta_{ki}, \quad L_k(1) = 1 \quad \text{and} \quad L_k(-1) = (-1)^k,$$

we have

$$\int_{-1}^{1} \phi_i^N(t)\dot{\epsilon}^N(t)dt = \begin{cases} \sum\limits_{j=N+1}^{\infty} a_j(-1)^{j+1}, & i = 0, \\ \sum\limits_{j=N+1}^{\infty} a_j, & i = N, \\ 0, & i \neq 0, N. \end{cases}$$

Remark 3.6. *The weak integral approximation to the derivative $\dot{x}(t)$ follows from Remark 3.5. Let*

$$x^N(t) = \sum_{j=0}^{N} a_j L_j(t),$$
$$\dot{x}^N(t) = \sum_{j=0}^{N} a_j \dot{L}_j(t).$$

Multiplying $\dot{x}$ by a test function $\phi^N(t)$ then integrating over the domain gives

$$\int_{-1}^{1} \phi_i^N(t)\dot{x}(t)dt = \int_{-1}^{1} \phi_i^N(t)\dot{x}^N(t)dt + \int_{-1}^{1} \phi_i^N(t)\dot{\epsilon}^N(t)dt,$$

for $i = 0, 1, \dots, N$, where the residual term, $\epsilon^N(t)$, and its derivative, $\dot{\epsilon}^N(t)$, can be expressed by

$$\epsilon^N(t) = \sum_{j=N+1}^{\infty} a_j L_j(t),$$
$$\dot{\epsilon}^N(t) = \sum_{j=N+1}^{\infty} a_j \dot{L}_j(t).$$

From Remark 3.5, the weak integral residual term is

$$\int_{-1}^{1} \phi_i^N(t)\dot{\epsilon}^N(t)dt = \sum_{j=N+1}^{\infty} a_j \int_{-1}^{1} \phi_i^N(t)\dot{L}_j(t)dt = \begin{cases} \sum\limits_{j=N+1}^{\infty} a_j(-1)^{j+1}, & i = 0, \\ \sum\limits_{j=N+1}^{\infty} a_j, & i = N, \\ 0, & i \neq 0, N. \end{cases}$$

Therefore, the error in the weak integral differentiation term may be bounded as

$$\left|\int_{-1}^{1} \phi_i^N(t)\dot{x}(t)dt - \int_{-1}^{1} \phi_i^N(t)\dot{x}^N(t)dt\right| \leq \begin{cases} \sum\limits_{j=N+1}^{\infty} |a_j|, & i = 0, \\ \sum\limits_{j=N+1}^{\infty} |a_j|, & i = N, \\ 0, & i \neq 0, N. \end{cases}$$

In other words, the accuracy of the weak integral approximation to $\dot{x}$ is related to the Legendre spectral coefficients of the dropped modes. If the Legendre spectral coefficients of the dropped modes are negligible, that is

$$\sum_{j=N+1}^{\infty} |a_j| \leq \delta,$$

where $\delta \ll 1$, then

$$\int_{-1}^{1} \phi_i^N(t)\dot{x}(t)dt \approx \int_{-1}^{1} \phi_i^N(t)\dot{x}^N(t)dt,$$

for $i = 0, 1, \ldots, N$. The weak integral approximation to $\dot{x}$ will be similar to the accuracy of the approximation to x itself since

$$\left\|\epsilon^N(t)\right\|_{L^\infty} = \left\|\sum_{j=N+1}^{\infty} a_j L_j(t)\right\|_{L^\infty} \leq \sum_{j=N+1}^{\infty} |a_j| \|L_j(t)\|_{L^\infty} \leq \sum_{j=N+1}^{\infty} |a_j|.$$

Jackson's Theorem (Lemma 3.1) along with Remark 3.6 present persuasive arguments for the use of the weak integral approximation (a.k.a. Galerkin methods) in place of traditional collocation techniques for approximating system dynamics in direct methods for optimal control. We will see that Galerkin methods may be formulated to efficiently solve multi-scale problems (see Section 5.4). As a preview, Figure 15 shows a comparison of the exact solution to Example 3.1 with the multi-scale Galerkin optimal control formulation numerical solutions with $N_{x_1} = 3$, $N_{x_2} = 43$ and $N_u = 1$. This problem was solved with optimality and feasibility tolerances of 5×10^{-4} and 5×10^{-3}, respectively, and the exact solution was used as an initial guess.

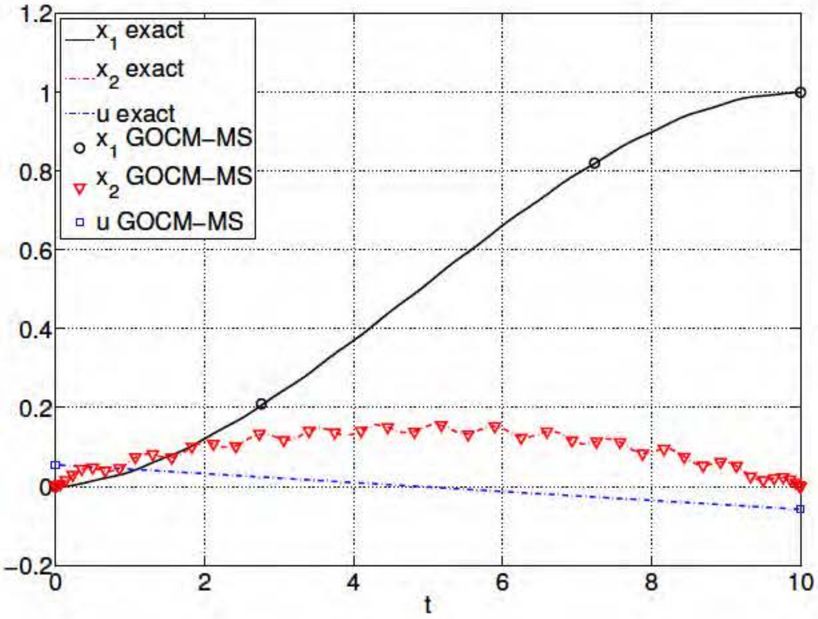


Figure 15: Exact solution and GOCM-MS approximation with $N_{x_1} = 3$, $N_{x_2} = 43$ and $N_u = 1$ for Example 3.1.

Simulations show that the multi-scale Legendre PS formulation becomes infeasible for the multi-scale approximation with orders $N_{x_1} = 3$, $N_{x_2} = 43$ and $N_u = 1$ for any reasonable set of optimality and feasibility tolerances selected.

Although the multi-scale Galerkin optimal control formulation shows promise in solving multi-scale problems, the advantages of the weak integral form are not limited to this problem set. Additionally, Galerkin formulations allow for the weak imposition of boundary conditions. That is, end conditions may be enforced only up to the accuracy of the approximation itself. Remark 2.9 highlights this property. Galerkin formulations with weak enforcement of boundary conditions have been shown to produce improved accuracies in many applications. A detailed discussion is given by Canuto et al. (see Section 3.7 of [46]). The Galerkin formulation with weak imposition of end conditions may also allow for problem discretizations with other than LGL points, such as LGR and LG (see

Sections 5.5.2.1 and 5.5.2.2, respectively). An important highlight of the Galerkin formulations is that the feasibility and consistency theorems are proved for problems with continuous and/or *piecewise continuous* controls (depending on the Galerkin formulation).

Lastly, Galerkin methods, as shown in Section 2.3.2, may be easily formulated as element-based methods, both continuous and discontinuous (see Sections 5.2 and 5.3, respectively). These element-based formulations may have benefits in approximating solutions to optimal control problems with multiple stages or those with discontinuous solutions, such as bang-bang control problems. As compared to global methods, these element-based techniques may be formulated to require less computational effort and memory. Additionally, the discontinuous Galerkin formulation may advantage from parallel computing. Chapter 4 will introduce a new numerical technique for solving nonlinear optimal control problems founded upon the Galerkin methods outlined in Section 2.3.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 4:
GENERAL GALERKIN OPTIMAL CONTROL FORMULATION

There are four parts to the numerical solution to an optimal control problem using the direct method: discretization of the system dynamics, discretization of the state-control constraints, integration of the cost function and solving the NLP. In the Galerkin optimal control approach introduced in [73, 74], we use Galerkin techniques to discretize the system dynamics based on LGL quadrature nodes. Recall, that the LGL points, $\{t_j\}_{j=0}^{N}$, are the roots of Equation (2.42) and therefore include the endpoints, $t = \pm 1$. Thus the discretization works in the interval of $[-1, 1]$ and will then provide the framework for our problem (e.g., the state-control constraints will be discretized at these nodes). Recall that LGL quadrature rule will provide zero error for polynomial integrands of less than or equal to $2N - 1$ [45]. Finally, LGL quadrature rule will be used to integrate the cost function. The resulting optimization problem can be solved using existing NLP algorithms.

In addition to the general Galerkin optimal control formulation, this chapter contains a number of important feasibility and consistency results. Theorems 4.1 and 4.2 prove that nonlinear program Problems GOCM-$\tilde{\text{S}}$ and GOCM-S (presented in Section 4.2) have feasible solutions to Problem B, where controls may be *piecewise continuous*. Additionally, Theorems 4.3 and 4.4 prove that the general Galerkin optimal control numerical approximation is consistent. That is, nonlinear programming Problems GOCM-$\tilde{\text{S}}$ and GOCM-S are consistent approximations to the continuous optimal control Problem B.

4.1. Method for Approximation

In the general Galerkin optimal control approximation to Problem B, the state trajectory, $x(t)$, is approximated with globally interpolating N-th order Lagrange polynomials, $\{\phi_j^N\}_{j=0}^N$, defined on a grid of LGL nodes, $\{t_j\}_{j=0}^N$,

$$x(t) \approx x^N(t) = \sum_{j=0}^{N} \phi_j^N(t)\bar{x}^{Nj}.$$

Due to the property of the Lagrange polynomials, $\phi_j^N(t_i) = \delta_{ij}$, we have

$$\bar{x}^{Nj} = x^N(t_j), \quad j = 0, 1, \ldots, N.$$

Also, let $u^N(t)$ be an interpolating function of $\{\bar{u}^{Nj}\}_{j=0}^N$,

$$u^N(t) = \sum_{j=0}^{N} \psi_j^N(t)\bar{u}^{Nj},$$

where $\{\psi_j^N\}_{j=0}^N$ is any set of continuous functions (not necessarily polynomials) with the property $\psi_j^N(t_i) = \delta_{ij}$. In the general Galerkin optimal control approach, a solution to the differential equation $\dot{x} - f(x, u) = 0$ may be approximated at the LGL nodes with the following weak integral formulation [44]

$$\int_{-1}^{1} \phi_i^N(t) \left(\frac{dx^N(t)}{dt} - f(x^N(t), u^N(t)) \right) dt = 0, \tag{4.1}$$

for $i = 0, 1, \ldots, N$. In terms of the approximating polynomials, the system of equations becomes

$$\sum_{j=0}^{N} \int_{-1}^{1} \phi_i^N \frac{d\phi_j^N}{dt} dt \, \bar{x}^{Nj} - \int_{-1}^{1} \phi_i^N f(x^N, u^N) dt = 0,$$

for $i = 0, 1, \ldots, N$, and in matrix-vector form is given by

$$\sum_{j=0}^{N} D_{ij}\bar{x}^{Nj} - c_i = 0, \quad i = 0, 1, \ldots, N.$$

The $(N+1) \times (N+1)$ Galerkin differentiation matrix, D, is defined by

$$D_{ij} = \int_{-1}^{1} \phi_i^N(t) \frac{d\phi_j^N(t)}{dt} dt, \quad i, j = 0, 1, \ldots, N, \tag{4.2}$$

and the $(N+1) \times 1$ right-hand-side (RHS) vector, c, is defined as

$$c_i = \int_{-1}^{1} \phi_i^N(t) f(x^N(t), u^N(t)) dt, \quad i = 0, 1, \ldots, N,$$

The Lagrange polynomials, $\{\phi_i^N\}_{i=0}^N$, and their derivatives, $\{\dot{\phi}_j^N\}_{j=0}^N$, are given by definitions (2.32) and (2.38), respectively. If LGL quadrature rule is used with $Q = N$ quadrature points, the differentiation matrix, D, can be calculated exactly by the relationship

$$D_{ij} = \sum_{k=0}^{Q} \phi_i^N(t_k) \frac{d\phi_j^N}{dt}(t_k) w_k = \frac{d\phi_j^N}{dt}(t_i) w_i = A_{ij} w_i, \quad i, j = 0, 1, \ldots, N, \tag{4.3}$$

where the LGL weights, $\{w_i\}_{i=0}^N$, are defined by Equation (2.49) and A is the Legendre PS differentiation matrix (2.57). The RHS vector, c, may also be approximated with quadrature with the relationship

$$c_i \approx \sum_{k=0}^{Q} \phi_i^N(t_k) f(x^N(t_k), u^N(t_k)) w_k, \tag{4.4}$$

for $i = 0, 1, \ldots, N$. If again, LGL quadrature rule is used with $Q = N$ quadrature points, the RHS vector approximation, $\bar{c}^N$, may be simplified as

$$\bar{c}^{Ni} = \sum_{k=0}^{N} \phi_i^N(t_k) f(x^N(t_k), u^N(t_k)) w_k = f(\bar{x}^{Ni}, \bar{u}^{Ni}) w_i, \quad i = 0, 1, \ldots, N. \tag{4.5}$$

Remark 4.1. *Recall that for LGL quadrature rule, integration is exact for polynomial integrands of degree less than or equal to $2N-1$. If $Q = (N+1)$ integration points are used, the RHS vector will integrate exactly when $f(x(t), u(t))$ is linear in $x(t)$ and $u(t)$. In the case of a nonlinear function f, the accuracy of integration (and therefore the accuracy* *of the overall approximation) can be improved by increasing the number of quadrature points Q. However, in most cases, increasing the accuracy of integration by increasing Q will significantly add to computation time due to the required interpolation of the state and control vectors. This will be discussed in greater detail in Section 5.5.1.*

Therefore system (4.1) may be simplified into the form

$$\sum_{j=0}^{N} D_{ij}\bar{x}^{Nj} - \bar{c}^{Ni} = 0, \quad i = 0, 1, \ldots, N. \tag{4.6}$$

Remark 4.2. *In form (4.6), the resulting Galerkin equations that must be satisfied are*

$$\left(\sum_{j=0}^{N} A_{ij}\bar{x}^{Nj} - f(\bar{x}^{Ni}, \bar{u}^{Ni})\right) w_i = 0, \tag{4.7}$$

for $i = 0, 1, \ldots, N$. This implies the following,

$$\sum_{j=0}^{N} A_{ij}\bar{x}^{Nj} - f(\bar{x}^{Ni}, \bar{u}^{Ni}) = 0, \tag{4.8}$$

for $i = 0, 1, \ldots, N$. Note that (4.8) is the same set of equations that would be relaxed when using the Legendre PS method (Section 3.2). Hence, numerical solutions to system (4.8) found via the collocation method will satisfy the Galerkin relationships in (4.6). However, inequality versions of (4.7) and (4.8) are used for computational purposes. As suggested by Jackson's Theorem, a solution of the inequality version of (4.7) may not satisfy (4.8). In fact, the analysis for the Galerkin numerical formulation is based on the L^2-norm. As a result, the inequality bound for the Galerkin formulation is not simply a multiple of the quadrature weight. Shown in Equation (4.13), the upper bound of the inequality includes a factor of $\sqrt{w}$. However, this relationship draws a clear connection between the general Galerkin optimal control formulation and the Legendre PS method, and will be exploited in the proof of convergence (Theorem 4.4).

Remark 4.3. *Due to the results of Gong et al.* [5]*, we know a feasible solution to the equality dynamical constraint may not exist. In order to guarantee feasibility of the discretized problem a relaxation of this constraint is used.*

The dynamical constraint becomes

$$\left\| \sum_{j=0}^{N} D_{ij}\bar{x}^{Nj} - \bar{c}^{Ni} \right\|_{\infty} \leq \delta^{N}, \quad i = 0, 1, \ldots, N,$$

where δ^N is the feasibility tolerance. The endpoint conditions and path constraints are approximated similarly by

$$\begin{aligned} \left\| e(\bar{x}^{N0}, \bar{x}^{NN}) \right\|_{\infty} &\leq \delta^{N}, \\ h(\bar{x}^{Ni}, \bar{u}^{Ni}) &\leq \delta^{N} \cdot \mathbf{1}, \quad i = 0, 1, \ldots, N, \end{aligned}$$

where $\mathbf{1}$ denotes $[1, \ldots, 1]^T$. Lastly, the cost functional $J[x(\cdot), u(\cdot)]$ is approximated by the LGL quadrature rule,

$$J[x(\cdot), u(\cdot)] \approx \bar{J}^{N}(\bar{x}^{N}, \bar{u}^{N}) = \sum_{i=0}^{N} F(\bar{x}^{Ni}, \bar{u}^{Ni}) w_i + E(\bar{x}^{N0}, \bar{x}^{NN}),$$

where $\bar{x}^N = \left[\bar{x}^{N0},\ \bar{x}^{N1}, \ldots, \bar{x}^{NN}\right]$ and $\bar{u}^N = \left[\bar{u}^{N0},\ \bar{u}^{N1}, \ldots, \bar{u}^{NN}\right]$. To allow for a practical search area for the optimal solution the following constraints are added

$$\{\bar{x}^{Ni} \in \boldsymbol{X}, \bar{u}^{Ni} \in \boldsymbol{U},\ i = 0, 1, \ldots, N\},$$

where $\boldsymbol{X}$ and $\boldsymbol{U}$ are the search regions that contain the optimal solution of the discretized nonlinear optimization.

4.2. Computation Strategy

The computation strategy for the Galerkin optimal control formulation with strong enforcement of BCs is presented in two forms. First, the strategy for the continuous problem, in terms of the approximating polynomials is outlined, denoted as GOCM-$\tilde{\text{S}}$. Next, the discrete problem, discretized on a LGL grid is presented, denoted as GOCM-S.

Definition 4.1. Function $g(t)$ is called piecewise C^1 if $\exists\ t_0 = -1 < t_1 < \cdots < t_k = 1$ such that $g(t)$ is C^1 on each subinterval (t_i, t_{i+1}), $i = 0, \ldots, k-1$; $\lim\limits_{t \to t_0^+} g(t)$, $\lim\limits_{t \to t_k^-} g(t)$ and $\lim\limits_{t \to t_i^{+/-}} g(t)$ exist for $i = 1, \ldots, k-1$; and $g(t)$ is either left or right continuous at each point t_i.

4.2.1. Computation Strategy for GOCM-$\tilde{\text{S}}$

The computational strategy of the GOCM-$\tilde{\text{S}}$ is to find the feasible solution $x^N(t) \in \boldsymbol{X}$ and $u^N(t) \in \boldsymbol{U}$ for the following cases:

Case 1. $u(\cdot)$ is piecewise C^0 and $x(\cdot) \in C^0$ and piecewise C^1,

Case 2. $u(\cdot),\ \dot{x}(\cdot) \in H^{m-1}$ (see Appendix C) and $m \geq 2$,

that minimizes

$$J(x^N(\cdot), u^N(\cdot)) = \int_{-1}^{1} F(x^N(t), u^N(t))dt + E(x^N(-1), x^N(1)), \tag{4.9}$$

subject to the Galerkin constraints

$$\begin{aligned} \left\| \int_{-1}^{1} \phi_i^N(t) \left(\dot{x}^N(t) - f(x^N(t), u^N(t)) \right) dt \right\|_\infty &\leq M w_i^{\frac{1}{2}} N^{-\alpha}, \quad i = 0, 1, \ldots, N, \\ \left\| e(x^N(-1), x^N(1)) \right\|_\infty &\leq M N^{-\alpha}, \\ \left\| h^+(x^N(t), u^N(t)) \right\|_{L^2} &\leq M N^{-\alpha}, \end{aligned} \tag{4.10}$$

where $\alpha = \frac{1}{2}$ and $(m-1)$, for Case 1 and 2, respectively; M is a constant independent of N and

$$h^+ = \begin{cases} h, & h > 0, \\ 0, & h \leq 0. \end{cases} \tag{4.11}$$

4.2.2. Computation Strategy for GOCM-S

The computational strategy of the GOCM-S is to find the feasible solution $\bar{x}^N \in \boldsymbol{X}$ and $\bar{u}^N \in \boldsymbol{U}$ for the following cases:

Case 1. $u(\cdot)$ is piecewise C^0 and $x(\cdot) \in C^0$ and piecewise C^1,

Case 2. $u(\cdot),\ \dot{x}(\cdot) \in H^{m-1}$ (see Appendix C), $m \geq 2$ and $\dot{x}^{(m-1)}(t)$ is of bounded variation in $t \in [-1, 1]$ (see Appendix A),

that minimizes

$$\bar{J}^N(\bar{x}^N, \bar{u}^N) = \sum_{i=0}^{N} F(\bar{x}^{Ni}, \bar{u}^{Ni}) w_i + E(\bar{x}^{N0}, \bar{x}^{NN}), \tag{4.12}$$

subject to the Galerkin constraints

$$\begin{aligned} \left\| \sum_{j=0}^{N} D_{ij} \bar{x}^{Nj} - \bar{c}^{Ni} \right\|_\infty &\leq M w_i^{\frac{1}{2}} N^{-\alpha}, \quad i = 0, 1, \ldots, N, \\ \left\| e(\bar{x}^{N0}, \bar{x}^{NN}) \right\|_\infty &\leq M N^{-\alpha}, \\ h(\bar{x}^{Ni}, \bar{u}^{Ni}) &\leq M N^{-\alpha} \cdot \mathbf{1}, \quad i = 0, 1, \ldots, N, \end{aligned} \tag{4.13}$$

where $\alpha = \frac{1}{2}$ and $(m-1)$, for Case 1 and 2, respectively; and M is a constant independent of N.

4.3. Feasibility of Solutions

In order to guarantee feasibility of the discretization, Theorems 4.1 and 4.2 show that a relaxation of the dynamical equality constraint to inequality is required, for GOCM-

S̃ and GOCM-S, respectively. However, first a buildup of lemmas are required for each theorem.

Lemma 4.1. [46] *Let $p^N(t)$ be the N-th order truncated Legendre series polynomial approximation to $\zeta \in H^m$, $t \in [-1,1]$, then*

$$\left\|\zeta(t) - p^N(t)\right\|_{L^2} \leq a_1 a_0 N^{-m}, \quad \forall\, t \in [-1,1]$$

and

$$\left\|\zeta(t) - p^N(t)\right\|_{L^\infty} \leq a_3 a_2 N^{\frac{1}{2}-m}, \quad \forall\, t \in [-1,1],$$

where a_1 and a_3 are constants independent of N; $a_0 = |\zeta|_{H^{m;N}}$, the Sobolev seminorm of ζ (see Appendix C); $a_2 = V(\zeta^{(m)})$, the total variation of $\zeta^{(m)}$ (see Appendix A); and $m \geq 0$. ($p^N(t)$with the smallest norm $\left\|\zeta(t) - p^N(t)\right\|_{L^2}$ is called the N-th order best polynomial approximation of ζ in the L^2-norm.)

Lemma 4.2. *Let $\zeta(t) = g(t) + hu_{t_c}(t)$, $t \in [-1,1]$, where ζ, $u_{t_c} \in H^0$, $g \in H^1$, and $u_{t_c}(t) = u(t - t_c)$ is the unit step function defined by*

$$u_{t_c}(t) = \begin{cases} 0, & -1 \leq t < t_c, \\ 1, & t_c \leq t \leq 1. \end{cases}$$

Also, let $p^N(t) = \sum\limits_{n=0}^{N} \hat{p}_n L_n$ be the N-th order truncated Legendre series polynomial approximation to ζ. Then

$$\left\|\zeta(t) - p^N(t)\right\|_{L^2} < b_1 b_0 N^{-1} + b_2(t_0, h) N^{-\frac{1}{2}}, \quad \forall\, t \in [-1,1],\ t_c \neq -1, 1 \text{ and } |h| < \infty,$$

where b_1 and b_2 are constants independent of N, and $b_0 = \|g\|_{H^1}$. ($p^N(t)$ with the smallest norm $\left\|\zeta(t) - p^N(t)\right\|_{L^2}$ is called the N-th order best polynomial approximation of ζ in the L^2-norm.)

Proof. Let

$$g^N(t) = \sum_{n=0}^{N} g_n L_n \quad \text{and} \quad u_{t_c}^N(t) = \sum_{n=0}^{N} u_n L_n$$

be the truncated Legendre series of g and u_{t_c}, respectively. Then

$$\begin{aligned} p^N(t) = \sum_{n=0}^{N} p_n L_n &= g^N(t) + h u_{t_c}(t) \\ &= \sum_{n=0}^{N} g_n L_n + h \sum_{n=0}^{N} u_n L_n, \end{aligned}$$

for $t \in [-1, 1]$, where

$$p_n = g_n + h u_n, \quad n = 0, \dots, N.$$

Therefore,

$$\begin{aligned} \left\| \zeta(t) - p^N(t) \right\|_{L^2} =& \left\| \zeta(t) - \sum_{n=0}^{N} p_n L_n \right\|_{L^2} \\ =& \left\| \left(g(t) - g^N(t) \right) + h \left(u_{t_c}(t) - u_{t_c}^N(t) \right) \right\|_{L^2} \\ \leq& \left\| g(t) - g^N(t) \right\|_{L^2} + |h| \left\| u_{t_c}(t) - u_{t_c}^N(t) \right\|_{L^2} \\ =& \left\| g(t) - \sum_{n=0}^{N} g_n L_n \right\|_{L^2} + |h| \left\| u_{t_c}(t) - \sum_{n=0}^{N} u_n L_n \right\|_{L^2}. \end{aligned}$$

From Lemma 4.1,

$$\left\| g(t) - \sum_{n=0}^{N} g_n L_n \right\|_{L^2} \leq b_1 b_0 N^{-1}, \quad t \in [-1, 1],$$

where g^N is called the polynomial of best approximation of g in the L^2-norm, b_1 is a constant independent of N and $b_0 = \|g\|_{H^1}$. Additionally, from [45],

$$|h|\left\|u_{t_c}(t) - \sum_{n=0}^{N} u_n L_n\right\|_{L^2} \leq |h| \left(\sum_{n=N+1}^{\infty} \gamma_n u_n^2\right)^{\frac{1}{2}}, \tag{4.14}$$

where the normalizing constants, $\{\gamma_n\}_{n=0}^{\infty}$, for the Legendre polynomials are given by Equation (2.27) and the Legendre expansion coefficients, $\{u_n\}_{n=0}^{\infty}$, are defined by Equation (2.26). Due to the properties of the Legendre polynomials,

$$L_n(t) = \frac{1}{2n+1}\left(L'_{n+1}(t) - L'_{n-1}(t)\right), \text{ and } L_n(1) = 1,$$

the Legendre coefficients have the relationship

$$u_n = \frac{1}{2}\int_{t_c}^{1} \left(L'_{n+1}(t) - L'_{n-1}(t)\right) dt = \frac{1}{2}\left(L_{n-1}(t_c) - L_{n+1}(t_c)\right),$$

or may be expressed by

$$|u_n| = \frac{1}{2}|(L_{n-1}(t_c) - L_{n+1}(t_c))| \leq \frac{1}{2}\left(|L_{n-1}(t_c)| + |L_{n+1}(t_c)|\right).$$

Since the Legendre polynomial has the bound [72]

$$|L_n(t)| < \frac{4(\frac{2}{\pi})^{\frac{1}{2}}}{n^{\frac{1}{2}}(1-t^2)^{\frac{1}{4}}}, \quad t \neq -1, 1,$$

we have the following bound on u_n,

$$|u_n| < \frac{2(\frac{2}{\pi})^{\frac{1}{2}}}{|h| n^{\frac{1}{2}}(1-t_c^2)^{\frac{1}{4}}}\left(\frac{1}{(n-1)^{\frac{1}{2}}} + \frac{1}{(n+1)^{\frac{1}{2}}}\right) < \frac{b_2}{|h|}\frac{1}{n^{\frac{1}{2}}},$$

for $n \geq 2$, where

$$b_2(t_c, h) = \frac{8}{\pi^{\frac{1}{2}}} \frac{|h|}{(1 - t_c^2)^{\frac{1}{4}}}, \quad t_c \neq -1, 1.$$

This is since

$$\frac{1}{(n-1)^{\frac{1}{2}}} + \frac{1}{(n+1)^{\frac{1}{2}}} < \frac{2}{(n-1)^{\frac{1}{2}}} \leq \frac{2\sqrt{2}}{n^{\frac{1}{2}}}, \quad n \geq 2.$$

Therefore, Equation (4.14) has the bound,

$$|h| \left\| u_{t_c}(t) - \sum_{n=0}^{N} \hat{u}_n L_n \right\|_{L^2} < b_2 \left(\sum_{n=N+1}^{\infty} \frac{1}{(n + \frac{1}{2})} \frac{1}{n} \right)^{\frac{1}{2}} < b_2 \left(\sum_{n=N+1}^{\infty} \frac{1}{n^2} \right)^{\frac{1}{2}}.$$

However, from the Integral Test Estimate [75],

$$\lim_{b \to \infty} \int_{N+1}^{b} \frac{1}{x^2} dx \leq \sum_{n=N+1}^{\infty} \frac{1}{n^2} \leq \lim_{b \to \infty} \int_{N}^{b} \frac{1}{x^2} dx = \frac{1}{N}.$$

Hence,

$$|h| \left\| u_{t_c}(t) - \sum_{i=0}^{N} \hat{u}_n L_n \right\|_{L^2} < b_2 N^{-\frac{1}{2}},$$

and finally,

$$\left\| \zeta(t) - p^N(t) \right\|_{L^2} < b_1 b_0 N^{-1} + b_2 N^{-\frac{1}{2}}, \quad \forall \, t \in [-1, 1], \; t_0 \neq -1, 1 \text{ and } |h| < \infty.$$

□

Lemma 4.3 (Hölder's Inequality)**.** [76] *Let the Hölder conjugates p and q be real numbers with the property that*

$$\frac{1}{p} + \frac{1}{q} = 1,$$

where $p > 1$ and $q > 1$. Then for any arbitrary complex-valued sequences $x = \{\xi_k\}_{k=0}^N$ and $y = \{v_k\}_{k=0}^N$ the following property holds

$$\sum_{k=0}^{N} |\xi_k v_k| \leq \left(\sum_{k=0}^{N} |\xi_k|^p \right)^{\frac{1}{p}} \left(\sum_{k=0}^{N} |v_k|^q \right)^{\frac{1}{q}}.$$

Moreover, when extended to integrals, Hölder's inequality takes the form

$$\int_a^b |f(t)g(t)| dt \leq \left(\int_a^b |f(t)|^p dt \right)^{\frac{1}{p}} \left(\int_a^b |g(t)|^q dt \right)^{\frac{1}{q}},$$

where f and g are assumed to be p-th and q-th power summable, respectively, on $t \in [a, b]$.

Lemma 4.4. *Let $\{\phi_i^N(t)\}_{i=0}^N$ be the Lagrange interpolating polynomials of order, N, defined on LGL grid $\{t_i\}_{i=0}^N$. Then, there exists a positive integer, N_0, such that, for any $N \geq N_0$,*

$$\left\| \phi_i^N \right\|_{L^2} \leq p w_i^{\frac{1}{2}} \leq q N^{-\frac{1}{2}},$$

for each $i = 0, 1, \ldots, N$, where $\{w_i\}_{i=0}^N$ are the LGL quadrature weights associated with the LGL points, $\{t_i\}_{i=0}^N$, and p and q are positive constants independent of N.

Proof. From [46], the discrete norm, $\|\xi_N\|_N = \left(\sum_{k=0}^{N} |\xi_k|^2 w_k \right)^{\frac{1}{2}}$, has the property

$$\|\xi_N\|_{L^2} \leq p \|\xi_N\|_N,$$

for $\xi_N \in \mathrm{P}_N$, where p is a positive constant, independent of N. Since $\phi^N \in \mathrm{P}_N$, we have

$$\left\| \phi_i^N \right\|_{L^2} \leq p \left\| \phi_i^N \right\|_N,$$

for each $i = 0, 1, \ldots, N$. Furthermore, from the property of the Lagrange polynomial, $\phi_j^N(t_i) = \delta_{ij}$, we have

$$\left\|\phi_i^N\right\|_N = \left(\sum_{k=0}^{N}\left|\phi_i^N(t_k)\right|^2 w_k\right)^{\frac{1}{2}} = w_i^{\frac{1}{2}},$$

for $i = 0, 1, \ldots, N$. Also, from [46] we have, for $i = 1, 2, \ldots, N-1$,

$$\frac{c_1}{N}(1-(t_i)^2)^{\frac{1}{2}} \leq w_i \leq \frac{c_2}{N}(1-(t_i)^2)^{\frac{1}{2}},$$

for constants, $0 < c_1 < c_2$, independent of i and N; for $i = 0, N$, we have

$$w_0, w_N = \frac{2}{N(N+1)}.$$

Therefore,

$$|w_i| \leq \frac{c_2}{N},$$

for each $i = 0, 1, \ldots, N$. Finally,

$$\left\|\phi_i^N\right\|_{L^2} \leq p w_i^{\frac{1}{2}} \leq q N^{-\frac{1}{2}}$$

holds for all $N > N_0$, where q is a constant independent of i and N. □

4.3.1. Feasibility of GOCM-S̃

Theorem 4.1 (Feasibility of GOCM-S̃)**.** *Given any feasible solution $t \mapsto (x, u)$, for Problem B, consider the following two cases:*

Case 1. $u(\cdot)$ is piecewise C^0 and $x(\cdot) \in C^0$ and piecewise C^1,

Case 2. $u(\cdot),\ \dot{x}(\cdot) \in H^{m-1}$ and $m \geq 2$.

Then, there exists a positive integer N_0 such that, for any $N \geq N_0$, GOCM-S̃ has a poly-

nomial feasible solution, $(x^N(t), u^N(t))$ *such that*

$$\left\|x(t) - x^N(t)\right\|_{L^2} \leq MN^{-\alpha} \quad \text{and} \quad \left\|u(t) - u^N(t)\right\|_{L^2} \leq MN^{-\alpha},$$

where $\alpha = \frac{1}{2}$ *and* $(m-1)$*, for Case 1 and 2, respectively; and* M *is a positive constant independent of* N*.*

Proof. Let $p(t)$ be the $(N-1)$-th order truncated Legendre polynomial approximation of $\dot{x}(t)$. By Lemmas 4.1 and 4.2 there is a constant c_0 independent of N, for any $N \geq N_0$, such that

$$\|\dot{x}(t) - p(t)\|_{L^2} \leq c_0 N^{-\alpha},$$

where $\alpha = \frac{1}{2}$ and $(m-1)$, for Case 1 and 2, respectively. Define

$$x^N(t) = \int_{-1}^{t} p(s)ds + x(-1).$$

Then $p(t) = \dot{x}^N(t)$ and

$$\left\|x(t) - x^N(t)\right\|_{L^2} \leq 2c_0 N^{-\alpha},$$

since, from Hölder's inequality (Lemma 4.3), we have

$$\begin{aligned} \left|x(t) - x^N(t)\right| &= \left|\int_{-1}^{t} (\dot{x}(s) - p(s))\, ds\right| \leq \int_{-1}^{t} |\dot{x}(s) - p(s)| ds \\ \leq& \sqrt{2} \left(\int_{-1}^{1} |\dot{x}(s) - p(s)|^2 ds\right)^{\frac{1}{2}} = \sqrt{2} \|\dot{x}(t) - p(t)\|_{L^2} \leq \sqrt{2} c_0 N^{-\alpha}. \end{aligned} \tag{4.15}$$

Let $u^N(t)$ be the N-th order Legendre polynomial so that

$$\left\|u(t) - u^N(t)\right\|_{L^2} \leq c_1 N^{-\alpha}.$$

From our Galerkin approximation, Hölder's inequality (Lemma 4.3), and Lemma 4.4, we have for each $k = 0, 1, \ldots, N$,

$$\begin{aligned}
&\left| \int_{-1}^{1} \phi_k^N(t) \left(\dot{x}^N(t) - f(x^N(t), u^N(t)) \right) dt \right| \\
&\leq \int_{-1}^{1} \left| \phi_k^N(t) \left(\dot{x}^N(t) - f(x^N(t), u^N(t)) \right) \right| dt \\
&\leq \left\| \phi_k^N(t) \right\|_{L^2} \left\| \dot{x}^N(t) - f(x^N(t), u^N(t)) \right\|_{L^2} \\
&\leq c_2 w_k^{\frac{1}{2}} \left\| \dot{x}^N(t) - f(x^N(t), u^N(t)) \right\|_{L^2} \\
&\leq c_2 w_k^{\frac{1}{2}} \left\| \dot{x}(t) - \dot{x}^N(t) \right\|_{L^2} + c_2 w_k^{\frac{1}{2}} \left\| \dot{x}(t) - f(x^N(t), u^N(t)) \right\|_{L^2} \\
&= c_2 w_k^{\frac{1}{2}} \| \dot{x}(t) - p(t) \|_{L^2} + c_2 w_k^{\frac{1}{2}} \left\| f(x(t), u(t)) - f(x^N(t), u^N(t)) \right\|_{L^2} \\
&= c_0 c_2 w_k^{\frac{1}{2}} N^{-\alpha} + c_2 l_1 w_k^{\frac{1}{2}} \left\| x(t) - x^N(t) \right\|_{L^2} + c_2 l_2 w_k^{\frac{1}{2}} \left\| u(t) - u^N(t) \right\|_{L^2} \\
&\leq c_0 c_2 w_k^{\frac{1}{2}} N^{-\alpha} + c_0 c_2 l_1 w_k^{\frac{1}{2}} N^{-\alpha} + c_1 c_2 l_2 w_k^{\frac{1}{2}} N^{-\alpha},
\end{aligned}$$

where $\{w_k\}_{k=0}^N$ are LGL quadrature weights and l_1 and l_2 are the Lipschitz constants of f with respect to x and u, respectively, which are independent of N. It follows that

$$\left| \int_{-1}^{1} \phi_k^N(t) \left(\dot{x}^N(t) - f(x^N(t), u^N(t)) \right) dt \right| \leq M w_k^{\frac{1}{2}} N^{-\alpha}$$

holds for each $k = 0, 1, \ldots, N$, and all $N > N_0$, where M is a constant independent of N.

For the endpoint condition we have

$$\begin{aligned}
\left| x(1) - x^N(1) \right| &= \left| \int_{-1}^{t} (\dot{x}(s) - p(s))\, ds \right| \leq \int_{-1}^{t} |\dot{x}(s) - p(s)| ds \\
\leq& \sqrt{2} \left(\int_{-1}^{1} |\dot{x}(s) - p(s)|^2 ds \right)^{\frac{1}{2}} = \sqrt{2} \| \dot{x}(t) - p(t) \|_{L^2} \leq \sqrt{2} c_0 N^{-\alpha},
\end{aligned}$$

so we have, by Lipschitz condition,

$$\left|e(x^N(-1), x^N(1))\right| \leq MN^{-\alpha}.$$

For the path constraint let $\mathcal{D} = \left\{t | h(x^N(t), u^N(t)) > 0\right\}, \overline{\mathcal{D}} = [-1, 1] \setminus \mathcal{D}$, since $h(x(t), u(t)) \leq 0$. Then

$$\begin{aligned}
\left\|h^+(x^N(t), u^N(t))\right\|_{L^2} &= \left(\int_{\mathcal{D}} (h(x^N(t), u^N(t)))^2 dt\right)^{\frac{1}{2}} \\
&\leq \left(\int_{\mathcal{D}} (h(x^N(t), u^N(t)) - h(x(t), u(t)))^2 dt\right)^{\frac{1}{2}} \\
&\leq \left(\int_{\mathcal{D}} (h(x^N(t), u^N(t)) - h(x(t), u(t)))^2 dt + \int_{\overline{\mathcal{D}}} (h(x^N(t), u^N(t)) - h(x(t), u(t)))^2 dt\right)^{\frac{1}{2}} \\
&= \left(\int_{-1}^{1} (h(x^N(t), u^N(t)) - h(x(t), u(t)))^2 dt\right)^{\frac{1}{2}} \\
&= \left\|h(x^N(t), u^N(t)) - h(x(t), u(t))\right\|_{L^2} \\
&\leq l_3 \left\|x(t) - x^N(t))\right\|_{L^2} + l_4 \left\|u(t) - u^N(t))\right\|_{L^2} \leq MN^{-\alpha},
\end{aligned}$$

where l_3 and l_4 are the Lipschitz constants of h with respect to x and u, respectively, which are independent of N. Hence

$$\left\|h^+(x^N(t), u^N(t))\right\|_{L^2} \leq MN^{-\alpha}.$$

Thus a solution $(x^N(t), u^N(t))$ to GOCM-S̃ is feasible! □

4.3.2. Feasibility of GOCM-S

Theorem 4.2 (Feasibility of GOCM-S). *Given any feasible solution $t \mapsto (x, u)$, for Problem B, consider the following two cases:*

Case 1. $u(\cdot)$ is piecewise C^0 and $x(\cdot) \in C^0$ and piecewise C^1,

Case 2. $u(\cdot)$, $\dot{x}(\cdot) \in H^{m-1}$ and $m \geq 2$.

Then, there exists a positive integer N_0 such that, for any $N \geq N_0$, GOCM-S has a feasible *solution,* $(\bar{x}^N, \bar{u}^N)$ *such that*

$$\left\|x(t) - x^N(t)\right\|_{L^2} \leq MN^{-\alpha},$$

where $\alpha = \frac{1}{2}$ *and* $(m-1)$*, for Case 1 and 2, respectively; and* M *is a positive constant independent of* N*. Additionally,* $u^N(t_i) = u(t_i)$*, for* $i = 0, 1, \ldots, N$*.*

Proof. Let $p(t)$be the $(N-1)$-th order truncated Legendre polynomial approximation of $\dot{x}(t)$ in the L^2-norm. By Lemma 4.1 there is a constant d_1 independent of N, for any $N \geq N_1$, such that

$$\|\dot{x}(t) - p(t)\|_{L^\infty} \leq d_1 N^{-\beta},$$

where $\beta = \left(m - \frac{3}{2}\right)$, for Case 2. For Case 1, we refer to [77–79] which show the truncated Legendre approximation for discontinuous functions with jump discontinuity (such as the step function defined in Lemma 4.2) displays Gibbs phenomenon. However, the maximum amplitude of the overshoot has a finite limit; we conclude that for Case 1, $\beta = 0$. Also, by Lemma 4.2 there is a constant d_2 independent of N, for any $N \geq N_2$, such that

$$\|\dot{x}(t) - p(t)\|_{L^2} \leq d_2 N^{-\alpha},$$

where $\alpha = \frac{1}{2}$ and $(m-1)$, for Case 1 and 2, respectively. Define

$$x^N(t) = \int_{-1}^{t} p(s)ds + x(-1).$$

Then $p(t) = \dot{x}^N(t)$ and

$$\left\|x(t) - x^N(t)\right\|_{L^2} \leq 2d_2 N^{-\alpha},$$

since, from Hölder's inequality (Lemma 4.3), we have

$$\left|x(t) - x^N(t)\right| = \left|\int_{-1}^{t} (\dot{x}(s) - p(s))\, ds\right| \leq \int_{-1}^{t} |\dot{x}(s) - p(s)| ds \tag{4.16}$$

$$\leq \sqrt{2} \left(\int_{-1}^{1} |\dot{x}(s) - p(s)|^2 ds\right)^{\frac{1}{2}} = \sqrt{2} \|\dot{x}(t) - p(t)\|_{L^2} \leq \sqrt{2} d_2 N^{-\alpha}. \tag{4.17}$$

Also, let $u^N(t)$ be an interpolating function of $u(t)$,

$$u^N(t) = \sum_{j=0}^{N} \psi_j^N(t) \bar{u}^{Nj},$$

where $\{\psi_j^N\}_{j=0}^N$ is any set of continuous functions (not necessarily polynomials) with the property $\psi_j^N(t_i) = \delta_{ij}$, and therefore

$$\bar{u}^{Nj} = u(t_j).$$

Since $x^N(t)$ is a N-th order polynomial, we have

$$\dot{x}^N(t_k) = \sum_{j=0}^{N} A_{kj} \bar{x}^{Nj},$$

where A is the $(N+1) \times (N+1)$ Legendre PS differentiation matrix (2.57) and

$$\bar{x}^{Nk} = x^N(t_k).$$

Recall that the LGL quadrature weights, $\{w_k\}_{k=0}^N$, have the property

$$w_k \leq d_3 N^{-1} (1 - (t_k)^2)^{\frac{1}{2}}, \quad k = 1, 2, \ldots, N-1,$$

for constant $d_3 > 0$ independent of k and N; for $k = 0, N$, we have

$$w_0, w_N = \frac{2}{N(N+1)}.$$

So, we have

$$\begin{aligned}
\left|\sum_{j=0}^{N} D_{kj}\bar{x}^{Nj} - \bar{c}^{Nk}\right| &= \left|\sum_{j=0}^{N} A_{kj}\bar{x}^{Nj} - f(\bar{x}^{Nk}, \bar{u}^{Nk})\right| w_k \\
= &\ \left|\dot{x}^N(t_k) - f(\bar{x}^{Nk}, \bar{u}^{Nk})\right| w_k = \left|p(t_k) - f(\bar{x}^{Nk}, \bar{u}^{Nk})\right| w_k \\
\leq &\ |p(t_k) - f(x(t_k), u(t_k))| w_k + \left|f(x(t_k), u(t_k) - f(\bar{x}^{Nk}, \bar{u}^{Nk})\right| w_k \\
= &\ |p(t_k) - \dot{x}(t_k)| w_k + \left|f(x(t_k), u(t_k) - f(\bar{x}^{Nk}, \bar{u}^{Nk})\right| w_k \\
&\leq \|p(t) - \dot{x}(t)\|_{L^\infty} w_k + l_1 \left|x(t_k) - x^N(t_k)\right| w_k \\
&\leq \ d_4 w_k N^{-\beta} + \sqrt{2} l_1 d_2 w_k N^{-\alpha},
\end{aligned}$$

for each $k = 0, 1, 2, \ldots, N$, where l_1 is the Lipschitz constants of f with respect to x. Putting this all together, we conclude that

$$\left|\sum_{j=0}^{N} D_{kj}\bar{x}^{Nj} - \bar{c}^{Nk}\right| \leq d_4 w_k N^{-\beta} + \sqrt{2} l_1 d_2 w_k N^{-\alpha} \leq M w_k^{\frac{1}{2}} N^{-\alpha},$$

for each $k = 0, 1, 2, \ldots, N$, and all $N > N_3$, where M is a constant independent of N.

For the endpoint condition, we have

$$\begin{aligned}
\left|x(t_N) - x^N(t_N)\right| = \left|x(1) - x^N(1)\right| &= \int_{-1}^{1} |\dot{x}(s) - p(s)| ds \\
\leq \sqrt{2} \left(\int_{-1}^{1} |\dot{x}(s) - p(s)|^2 ds\right)^{\frac{1}{2}} &= \sqrt{2} \|\dot{x}(t) - p(t)\|_{L^2} \leq \sqrt{2} d_2 N^{-\alpha}.
\end{aligned}$$

So, by Lipschitz condition,

$$\left|e(x^N(t_0), x^N(t_N))\right| \leq M N^{-\alpha}.$$

For the path constraint, the following estimate holds

$$\big|h(x(t),u(t)) - h(x^N(t),u^N(t))\big| \le l_2\big|x(t) - x^N(t)\big| \le \sqrt{2}l_2d_2N^{-\alpha},$$

for each $k = 0,1,2,\ldots,N$, where l_2 is the Lipschitz constants of h with respect to x. Hence,

$$h(\bar{x}^{Nk},\bar{u}^{Nk}) \le h(x(t_k),u(t_k)) + MN^{-\alpha}\cdot\mathbf{1}.$$

Thus a solution $(\bar{x}^N,\bar{u}^N)$ to GOCM-S is feasible! □

Remark 4.4. *Although, Theorems 4.1 and 4.2 do not provide exact feasibility tolerances for the existence of solutions to the GOCM-S̃ and GOCM-S, we can be confident that solutions do in fact exist. Precise bounds may be found experimentally, using a recursive refinement process, by increasing the order of the approximation, N, until all the constraints in the NLP are satisfied.*

4.4. Consistency of Solutions

Theorems 4.1 and 4.2 show that solutions exist to the GOCM-S̃ and GOCM-S, respectively. However, the question still remains—will these solutions converge to those that we seek? The answer is yes—Theorems 4.3 and 4.4 presented below show that solutions to GOCM-S̃ and GOCM-S, will in fact converge to the optimal solution of Problem B. However, first a definition and lemma are required.

4.4.1. Consistency of GOCM-S̃

Definition 4.2. The orthogonal system $\{\psi_k(t)\}_{k=0}^{\infty}$ is complete in L^2, $t \in [-1,1]$, if and only if, for $\xi(t) \in L^2$, the condition $\int_{-1}^{1}\psi_k(t)\xi(t)(t)dt = 0$, $\forall k \ge 0$, implies $\|\xi(t)\|_{L^2} = 0$.

Lemma 4.5. *Let $\epsilon^N(t) \in H^m$ with $m > \frac{1}{2}$ (see Appendix C). Assume*

$$\left| \int_{-1}^{1} \phi_i^N(t)\epsilon^N(t)dt \right| \leq b_0 w_i^{\frac{1}{2}} \delta^N,$$

for each $i = 0, 1, \ldots, N$, where $\{\phi_i^N(t)\}_{i=0}^N$ is the Lagrange interpolation polynomial of order N defined on LGL grid, $\{t_i\}_{i=0}^N$, $\{w_i\}_{i=0}^N$ are the associated LGL quadrature weights and b_0 is a constant independent of N. Also assume $\exists\ \epsilon(t) \in H^m$ with $m > \frac{1}{2}$, so that

$$\left\| \epsilon - \epsilon^N \right\|_{L^2} \leq b_1 \delta^N,$$

where $\delta^N \leq N^{-\alpha}, \alpha > \frac{1}{2}$ and b_1 is a constant independent of N. Then

$$\|\epsilon\|_{L^2} = 0.$$

Proof. Recall from Equation (2.35) that the orthogonal Legendre polynomials, $\{L_j(t)\}_{j=0}^N$, can be written as linear combinations of Lagrange polynomials, $\{\phi_i^N(t)\}_{i=0}^N$, defined on the LGL grid, $\{t_k\}_{k=0}^N$, as

$$L_j(t) = \sum_{i=0}^{N} L_j(t_i)\phi_i^N(t) = \sum_{i=0}^{N} V_{ij}^T \phi_i^N(t),$$

where V is the generalized Vandermonde matrix (2.30). From Hölder's inequality (Lemma 4.3) and Lemma 4.4, for each $i = 0, 1, \ldots, N$, we have

$$\begin{aligned}
\left|\int_{-1}^{1} \phi_i^N(t)\epsilon(t)dt\right| &= \left|\int_{-1}^{1} \phi_i^N(t)\left(\epsilon^N(t) - \epsilon^N(t) + \epsilon(t)\right) dt\right| \\
&\leq \left|\int_{-1}^{1} \phi_i^N(t)\left(\epsilon(t) - \epsilon^N(t)\right) dt\right| + \left|\int_{-1}^{1} \phi_i^N(t)\epsilon^N(t)dt\right| \\
&\leq \int_{-1}^{1} \left|\phi_i^N(t)\left(\epsilon(t) - \epsilon^N(t)\right)\right| dt + b_0 w_i^{\frac{1}{2}}\delta^N \\
&\leq \left\|\phi_i^N(t)\right\|_{L^2} \left\|\epsilon(t) - \epsilon^N(t)\right\|_{L^2} + b_0 w_i^{\frac{1}{2}}\delta^N \\
&\leq b_2 w_i^{\frac{1}{2}}\delta^N + b_0 w^{\frac{1}{2}}\delta^N = b_3 w_i^{\frac{1}{2}}\delta^N.
\end{aligned}$$

From [46], we have

$$\frac{b_4}{N}(1-(t_i)^2)^{\frac{1}{2}} \leq w_i \leq \frac{b_5}{N}(1-(t_i)^2)^{\frac{1}{2}}, \quad i = 1, 2, \ldots, N-1,$$

for constants $0 < b_4 < b_5$, independent of i and N; for $i = 0, N$, we have

$$w_0, w_N = \frac{2}{N(N+1)}.$$

Also recall, from [72], that

$$|L_j(t)| < \frac{4(\frac{2}{\pi})^{\frac{1}{2}}}{j^{\frac{1}{2}}(1-t^2)^{\frac{1}{4}}}, \quad t \neq -1, 1.$$

Since $\left|\int_{-1}^{1} \phi_i^N(t)\epsilon(t)dt\right| \le b_3 w_i^{\frac{1}{2}} \delta^N$, for each $i = 0, 1, \ldots, N$, we have

$$\begin{aligned}
\left|\int_{-1}^{1} L_j(t)\epsilon(t)dt\right| &= \left|\sum_{i=0}^{N} L_j(t_i) \int_{-1}^{1} \phi_i^N(t)\epsilon(t)dt\right| \\
&\le \sum_{i=0}^{N} |L_j(t_i)| \left|\int_{-1}^{1} \phi_i^N(t)\epsilon(t)dt\right| \le b_3 \sum_{i=0}^{N} |L_j(t_i)| w_i^{\frac{1}{2}} \delta^N \\
&\le \delta^N \left(b_3 \frac{2\sqrt{2}}{(N(N+1))^{\frac{1}{2}}} + b_6 \sum_{i=1}^{N-1} \left| \frac{1}{j^{\frac{1}{2}}(1-t_i^2)^{\frac{1}{4}}} \right| \left| \frac{(1-t_i^2)^{\frac{1}{4}}}{N^{\frac{1}{2}}} \right| \right) \\
&= \delta^N \left(b_2 \frac{2\sqrt{2}}{(N(N+1))^{\frac{1}{2}}} + b_6 \frac{N-1}{j^{\frac{1}{2}} N^{\frac{1}{2}}} \right) \le b_7 \delta^N \left(\frac{N}{j}\right)^{\frac{1}{2}},
\end{aligned}$$

for each $j = 1, \ldots, N$, and constant, b_7, for all $N \ge N_0$. Since $\delta^N \le N^{-\alpha}$ and $\alpha > \frac{1}{2}$, it follows that when $N \to \infty$ we have

$$\left|\int_{-1}^{1} L_j(t)\epsilon(t)dt\right| = 0,$$

for each $j = 0, 1, \ldots, N$. Since the Legendre polynomials, $\{L_j\}_{j=0}^{\infty}$, are complete in L^2 space [80] and $\epsilon \perp L_j$ for all $j = 0, 1, \ldots, N$, we can conclude,

$$\|\epsilon(t)\|_{L^2} = 0.$$

□

Theorem 4.3 (Consistency of GOCM-S̃)**.** *Suppose* $(x^N(t), u^N(t))$ *is a solution of GOCM-*$\tilde{S}$ *and there exists* $(x(t), u(t))$ *such that* $u(\cdot)$, $\dot{x}(\cdot) \in H^{m-1}$ *with* $m \ge 2$*. Also, suppose* $x^N(t) \to x(t)$ *uniformly, and*

$$\left\|x(t) - x^N(t)\right\|_{L^2} \le K\delta^N, \tag{4.18}$$

$$\left\|u(t) - u^N(t)\right\|_{L^2} \le K\delta^N, \tag{4.19}$$

where $\delta^N \leq N^{-\alpha}, \alpha > \frac{1}{2}$ and K is a constant independent of N. Then $(x(t), u(t))$ satisfies

$$\begin{cases} \|\dot{x}(t) - f(x(t), u(t))\|_{L^2} = 0, \\ e(x(-1), x(1)) = 0, \\ h(x(t), u(t)) \leq 0, \end{cases}$$

and is an optimal solution to Problem B.

Proof. Let $\epsilon^N(t) = \dot{x}^N(t) - f(x^N(t), u^N(t))$ and $\epsilon(t) = \dot{x}(t) - f(x(t), u(t))$. From Lemma 4.5, to prove $\|\epsilon(t)\|_{L^2} = 0$, it is enough to prove

$$\left| \int_{-1}^{1} \phi_k^N(t)\epsilon(t)dt \right| \leq c_0 w_k^{\frac{1}{2}} \delta^N,$$

for each $k = 0, 1, \ldots, N$, where $\delta^N \leq N^{-\alpha}$, $\alpha > \frac{1}{2}$. Consider

$$\begin{gathered} \left| \int_{-1}^{1} \phi_k^N(t)\epsilon(t)dt \right| \\ \leq \left| \int_{-1}^{1} \phi_k^N(t) \left(\dot{x}^N(t) - f(x^N(t), u^N(t)) \right) dt \right| + \left| \int_{-1}^{1} \phi_k^N(t) \left(\dot{x}(t) - \dot{x}^N(t) \right) dt \right| \\ + \left| \int_{-1}^{1} \phi_k^N(t) \left(f(x^N(t), u^N(t)) - f(x, u) \right) dt \right| \\ \leq c_2 w_k^{\frac{1}{2}} \delta^N + w_k^{\frac{1}{2}} \left\| \dot{x}(t) - \dot{x}^N(t) \right\|_{L^2} + w_k^{\frac{1}{2}} \left\| f(x(t), u(t)) - f(x^N(t), u(t)) \right\|_{L^2} \\ \leq c_2 w_k^{\frac{1}{2}} \delta^N + w_k^{\frac{1}{2}} \left\| x(t) - x^N(t) \right\|_{L^2} + l_1 w_k^{\frac{1}{2}} \left\| x(t) - x^N(t) \right\|_{L^2} + l_2 w_k^{\frac{1}{2}} \left\| u(t) - u^N(t) \right\|_{L^2} \\ \leq c_2 w_k^{\frac{1}{2}} \delta^N + K w_k^{\frac{1}{2}} \delta^N + l_1 K w_k^{\frac{1}{2}} \delta^N + l_2 K w_k^{\frac{1}{2}} \delta^N \leq c_0 w_k^{\frac{1}{2}} \delta^N, \end{gathered}$$

where l_1 and l_2 are the Lipschitz constants of f with respect to x and u, respectively, which are independent of N. It follows, from Lemma 4.5, that as $N \to \infty$ we have

$$\|\epsilon\|_{L^2} = \|\dot{x}(t) - f(x(t), u(t))\|_{L^2} = 0.$$

For the endpoint condition, since $x^N(t) \to x(t)$ uniformly, we have

$$x^N(1) \to x(1) \quad \text{and} \quad x^N(-1) \to x(-1).$$

Since, from the formulation of the computational strategy we have $\left|e(x^N(-1), x^N(1))\right| \leq MN^{-\alpha}$, we conclude that $e(x(-1), x(1)) = 0$ as $N \to \infty$.

For the path constraint, since $h(x(t), u(t))$ is piecewise C^1, if $h(x(t^*), u(t^*)) > 0$, $\exists$ an interval (a, b) in which $h(x(t), u(t)) > 0$. Then

$$\|h(x(t), u(t))\|_{L^2(a,b)} = \left(\int_a^b (h(x(t), u(t)))^2 dt\right)^{\frac{1}{2}} > 0. \tag{4.20}$$

However,

$$\begin{aligned}\|h(x(t), u(t))\|_{L^2(a,b)} &\leq \left\|h(x(t), u(t)) - h^+(x^N(t), u^N(t))\right\|_{L^2(a,b)} + \left\|h^+(x^N(t), u^N(t))\right\|_{L^2(a,b)} \\ &\leq \left\|h(x(t), u(t)) - h(x^N(t), u^N(t))\right\|_{L^2(a,b)} + MN^{-\alpha} \\ &\leq l_3\left\|x(t) - x^N(t))\right\|_{L^2} + l_4\left\|u(t) - u^N(t))\right\|_{L^2} + MN^{-\alpha},\end{aligned}$$

where l_3 and l_4 are the Lipschitz constants of h with respect to x and u, respectively, which are independent of N. Hence, this is a contradiction, therefore $h(x(t), u(t)) \leq 0$ as $N \to \infty$.

Suppose that $(x(t), u(t))$ is not optimal. Then $\exists$ $(x^*(t), u^*(t))$, so that

$$J\left(x^*(\cdot), u^*(\cdot)\right) < J\left(x(\cdot), u(\cdot)\right).$$

Also, $\exists$ $(x^*(t), u^*(t))$ such that

$$\left\|x^{*N}(t) - x^*(t)\right\|_{L^2} \leq MN^{-\alpha} \text{ and } \left\|u^{*N}(t) - u^*(t)\right\|_{L^2} \leq MN^{-\alpha},$$

where $(x^{*N}(t), u^{*N}(t))$ is a feasible trajectory of GOCM-S̃. Therefore

$$J\left(x^{*N}(\cdot), u^{*N}(\cdot)\right) \geq J\left(x^{N}(\cdot), u^{N}(\cdot)\right). \tag{4.21}$$

However,

$$\begin{aligned}
&\left|J\left(x^{N}(\cdot), u^{N}(\cdot)\right) - J\left(x(\cdot), u(\cdot)\right)\right| \\
\leq &\int_{-1}^{1}\left|F(x^N(t), u^N(t)) - F(x(t), u(t))\right|dt + \left|E(x^N(-1), x^N(1)) - E(x(-1), x(1))\right| \\
\leq &\sqrt{2}\left\|F(x^N(t), u^N(t)) - F(x(t), u(t))dt\right\|_{L^2} + \left|E(x^N(-1), x^N(1)) - E(x(-1), x(1))\right|.
\end{aligned}$$

Due to the Lipschitz condition and assumptions (4.18) and (4.19) we have

$$\lim_{N\to\infty}\left|J\left(x^{N}(\cdot), u^{N}(\cdot)\right) - J\left(x(\cdot), u(\cdot)\right)\right| = 0.$$

Similarly,

$$\lim_{N\to\infty}\left|J\left(x^{*N}(\cdot), u^{*N}(\cdot)\right) - J\left(x^{*}(\cdot), u^{*}(\cdot)\right)\right| = 0.$$

Therefore, from (4.21) we have

$$J\left(x^{*}(\cdot), u^{*}(\cdot)\right) \geq J\left(x(\cdot), u(\cdot)\right).$$

This is a contradiction, since we assumed

$$J\left(x^{*}(\cdot), u^{*}(\cdot)\right) < J\left(x(\cdot), u(\cdot)\right).$$

We conclude that $(x(t), u(t))$ achieves an optimal cost and therefore is an optimal solution to Problem B! □

4.4.2. Consistency of GOCM-S

Lemma 4.6 (Theorem 6.5.5, [81]). *Let f be Riemann integrable in $[-1,1]$ and $\{\phi_k^N(t)\}_{k=0}^N$ be the Lagrange polynomials of order N defined on a LGL grid, $\{t_k\}_{k=0}^N$. Define*

$$f^N(t) = \sum_{k=0}^{N} \phi_k^N(t) \bar{f}^{Nk},$$

where $\bar{f}^{Nk} = f^N(t_k)$, for $k = 0, 1, \ldots, N$. Then

$$\lim_{N\to\infty} \sum_{k=0}^{N} f(t_k) w_k = \lim_{N\to\infty} \int_{-1}^{1} f^N(t) dt = \int_{-1}^{1} f(t) dt,$$

where $\{w_k\}_{k=0}^N$ are LGL quadrature weights associated with the LGL points, $\{t_k\}_{k=0}^N$.

Theorem 4.4 (Consistency of GOCM-S). *Suppose $\left\{(\bar{x}^{Nk}, \bar{u}^{Nk}), 0 \le k \le N\right\}_{N=N_1}^{\infty}$ is a sequence of solutions to GOCM-S, $\left\{t \mapsto \left(x^N(t), u^N(t)\right)\right\}_{N=N_1}^{\infty}$ are their interpolating functions and there exists functions $(x(t), u(t))$ such that $u(\cdot),\ \dot{x}(\cdot) \in H^{m-1}$ with $m \ge 2$ and $\dot{x}^{(m-1)}(t)$ is of bounded variation in $t \in [-1,1]$ (see Appendix A). Also, suppose*

$$\lim_{N\to\infty} \left\|u(t) - u^N(t)\right\|_{L^\infty} = 0, \tag{4.22}$$

$$\lim_{N\to\infty} \left\|\dot{x}(t) - \dot{x}^N(t)\right\|_{L^\infty} = 0. \tag{4.23}$$

Then $(x(t), u(t))$ satisfies

$$\begin{cases} \|\dot{x}(t) - f(x(t), u(t))\|_{L^\infty} = 0, \\ e(x(-1), x(1)) = 0, \\ h(x(t), u(t)) \le 0, \end{cases}$$

and is an optimal solution to Problem B.

Proof. This proof follows the outline of Theorem 2 given by Gong et al. in [5]. First, from assumptions (4.22) and (4.23) it is easy to show

$$\lim_{N\to\infty}\left\|x(t)-x^N(t)\right\|_{L^\infty}=0.$$

Next, suppose that $(x(t),u(t))$ is not a solution. Then there is a time $\tau\in[-1,1]$ such that

$$\dot{x}(\tau)-f(x(\tau),u(\tau))\neq 0.$$

From [82], the LGL nodes, $\{t_i\}_{i=0}^N$, are dense when $N\to\infty$. Then there exists a sequence, $\{i^N\}$, where $0\le i^N\le N$, and with property

$$\lim_{N\to\infty}t_{i^N}=\tau,$$

so that

$$\dot{x}(\tau)-f(x(\tau),u(\tau))=\lim_{N\to\infty}\left(\dot{x}^N(t_{i^N})-f(x^N(t_{i^N}),u^N(t_{i^N}))\right)\neq 0. \tag{4.24}$$

Also, we have

$$\dot{x}^N(t_{i^N})=\sum_{j=0}^{N}A_{i^Nj}\bar{x}^{Nj},$$

where A is the Legendre PS differentiation matrix (2.57). Therefore, from Theorem 4.2,

$$\begin{aligned}\left|\sum_{j=0}^{N}D_{i^Nj}\bar{x}^{Nj}-\bar{c}^{Ni^N}\right|&=\left|\sum_{j=0}^{N}A_{i^Nj}\bar{x}^{Nj}-f(\bar{x}^{Ni^N},\bar{u}^{Ni^N})\right|w_{i^N}\\&=\left|\dot{x}^N(t_{i^N})-f(x^N(t_{i^N}),u^N(t_{i^N}))\right|w_{i^N}=Mw_{i^N}N^{\frac{3}{2}-m},\end{aligned}$$

where w_{i^N}, is the LGL quadrature weight associated with LGL point, t_{i^N}, for each i^N. This implies

$$\lim_{N\to\infty}\left(\dot{x}^N(t_{i^N}) - f(x^N(t_{i^N}), u^N(t_{i^N}))\right) = \lim_{N\to\infty} MN^{\frac{3}{2}-m} = 0,$$

which contradicts Equation (4.24). We conclude that $(x(t), u(t))$ is a solution.

For the path constraint, we consider the same contradiction argument given above.

For the endpoint condition, since $x^N(t) \to x(t)$ uniformly, we have

$$x^N(1) \to x(1) \quad \text{and} \quad x^N(-1) \to x(-1).$$

Since, from Theorem 4.2, we have

$$\left|e(x^N(-1), x^N(1))\right| \leq MN^{1-m},$$

we conclude that

$$e(x(-1), x(1)) = \lim_{N\to\infty} e(x^N(-1), x^N(1)) = \lim_{N\to\infty} e(\bar{x}^{N0}, x^{NN}) = 0.$$

For the cost functional we have

$$\bar{J}^N(\bar{x}^N, \bar{u}^N) = \sum_{k=0}^{N} F(\bar{x}^{Nk}, \bar{u}^{Nk})w_k + E(\bar{x}^{N0}, \bar{x}^{NN})$$

and

$$J(x(\cdot), u(\cdot)) = \int_{-1}^{1} F(x(t), u(t))dt + E(x(-1), x(1)).$$

By Lemma 4.6, we have

$$\int_{-1}^{1} F(x(t),u(t))dt = \lim_{N\to\infty}\sum_{k=0}^{N} F(x(t_k),u(t_k))w_k,$$

and therefore

$$\left|\int_{-1}^{1} F(x(t),u(t))dt\right| = \left|\lim_{N\to\infty}\left(\sum_{k=0}^{N} F(\bar{x}^{Nk},\bar{u}^{Nk})w_k + \sum_{k=0}^{N}\left[F(x(t_k),u(t_k) - F(\bar{x}^{Nk},\bar{u}^{Nk})\right] w_k\right)\right|.$$

However,

$$\begin{aligned}&\left|\lim_{N\to\infty}\sum_{k=0}^{N}\left[F(x(t_k),u(t_k)) - F(\bar{x}^{Nk},\bar{u}^{Nk})\right] w_k\right|\\ &\le \lim_{N\to\infty}\sum_{k=0}^{N}\left|F(x(t_k),u(t_k)) - F(x^N(t_k),u^N(t_k))\right|w_k\\ &\le l_1\lim_{N\to\infty}\sum_{k=0}^{N}\left|x(t_k) - x^N(t_k)\right|w_k + l_2\lim_{N\to\infty}\sum_{k=0}^{N}\left|u(t_k) - u^N(t_k)\right|w_k = 0,\end{aligned}$$

where l_1 and l_2 are the Lipschitz constants of F with respect to x and u. Thus we conclude,

$$\int_{-1}^{1} F(x(t),u(t))dt = \lim_{N\to\infty}\sum_{k=0}^{N} F(\bar{x}^{Nk},\bar{u}^{Nk})w_k.$$

Finally, by Lipschitz condition,

$$\lim_{N\to\infty} E(\bar{x}^{N0},\bar{x}^{NN}) = E(x(-1),x(1)),$$

and the limit

$$\lim_{N\to\infty}\bar{J}^N(\bar{x}^N,\bar{u}^N) = J\left(x(\cdot),u(\cdot)\right) \tag{4.25}$$

follows.

Let $\{(x^{*Nk}, u^{*Nk}), 0 \leq k \leq N\}_{N=N_1}^{\infty}$ be an optimal sequence of solutions to GOCM-S and $\{t \mapsto (x^{*N}(t), u^{*N}(t))\}_{N=N_1}^{\infty}$ be their interpolating functions. From Theorem 4.2,

$$\lim_{N\to\infty} \left\| u^*(t) - u^{*N}(t) \right\|_{L^\infty} = 0, \tag{4.26}$$

$$\lim_{N\to\infty} \left\| x^*(t) - x^{*N}(t) \right\|_{L^\infty} = 0, \tag{4.27}$$

and from (4.25) we have

$$J\left(x^*(\cdot), u^*(\cdot)\right) \leq J\left(x(\cdot), u(\cdot)\right) = \lim_{N\to\infty} \bar{J}^N(\bar{x}^N(\cdot), \bar{u}^N(\cdot)) \leq \lim_{N\to\infty} \bar{J}^N(x^{*N}(\cdot), u^{*N}(\cdot)).$$

Finally, from conditions (4.26) and (4.27) we conclude

$$J\left(x^*(\cdot), u^*(\cdot)\right) = J\left(x(\cdot), u(\cdot)\right).$$

Hence $(x(t), u(t))$ achieves an optimal cost and therefore is an optimal solution to Problem B! □

Remark 4.5. *Theorems 4.3 and 4.4 provide confidence that solutions not only existence to the GOCM-$\tilde{S}$ and GOCM-S, but solutions will converge to the optimal solution. However, questions still remain about the conditions under which assumptions (4.18), (4.19), (4.22) and (4.23) exist (as pointed out by Gong et al.* [5]*). Answers for similar questions have been provided for the Legendre PS method by Kang* [6]*, but like analysis for Galerkin optimal control is above the scope of this dissertation.*

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 5:
ALTERNATIVE FORMS OF GALERKIN OPTIMAL CONTROL

An advantage in using Galerkin optimal control is that it is a versatile family of formulations. There are a number of Galerkin forms that can be used to suit the problem at hand. In Chapter 4, the general formulation, Galerkin optimal control with strong enforcement of end conditions, was presented. This serves as the first of three global formulations that are outlined in this dissertation. The second global formulation is Galerkin optimal control with weak enforcement of boundary conditions, and will be discussed in Section 5.1. (Additionally, a third global Galerkin optimal control formulation with Legendre test functions will be presented in Chapter 6.) Important results in this chapter include Theorems 5.1 and 5.2, which prove that nonlinear program Problems GOCM-$\tilde{\text{W}}$ and GOCM-W (outlined in Section 5.1.2) have feasible solutions to Problem B, where the controls may be *piecewise continuous*.

Next, the element based formulations will be presented and are divided into two forms: Galerkin optimal control with element-based continuous and discontinuous Galerkin techniques, which will be presented in Sections 5.2 and 5.3, respectively. As alluded to in Chapter 3, the Galerkin weak integral form improves feasibility of the multi-scale approach highlighted. The method of approximation for the multi-scale Galerkin optimal control formulation will be outlined in Section 5.4. Finally, Section 5.5 will discuss modifications to the Galerkin optimal control formulations, such as over-integration of the RHS vector and the use of quadrature points other than LGL, such as LG and LGR points.

5.1. Galerkin optimal control with Weak Boundary Condition Enforcement

The general Galerkin optimal control strategy presented in Section 4.2 describes a formulation in which boundary conditions are enforced in a strong sense, via a constraint of the form

$$\left\| e(\bar{x}^{N0}, \bar{x}^{NN}) \right\|_\infty \leq \delta^N.$$

Recall, this boundary enforcement method is also incorporated into the Legendre PS method (see Section 3.2). This section presents an alternative formulation of Galerkin optimal control introduced in [73, 83], one that allows for enforcement of the problem end conditions in a weak sense through the dynamical constraint.

5.1.1. Method for Approximation

We now consider the Galerkin optimal control formulation with weak enforcement of boundary conditions. In this approximation to Problem B, the state trajectory, $x(t)$, is approximated with globally interpolating N-th order Lagrange polynomials,$\{\phi_j^N\}_{j=0}^N$, defined on a grid of LGL nodes, $\{t_j\}_{j=0}^N$,

$$x(t) \approx x^N(t) = \sum_{j=0}^N \phi_j^N(t) \bar{x}^{Nj}.$$

Due to the property of the Lagrange polynomials, $\phi_j^N(t_i) = \delta_{ij}$, we have

$$\bar{x}^{Nj} = x^N(t_j), \quad j = 0, 1, \ldots, N.$$

Also, let $u^N(t)$ be an interpolating function of $\{\bar{u}^{Nj}\}_{j=0}^N$,

$$u^N(t) = \sum_{j=0}^N \psi_j^N(t) \bar{u}^{Nj},$$

where $\{\psi_j^N\}_{j=0}^N$ is any set of continuous functions (not necessarily polynomials) with the property $\psi_j^N(t_i) = \delta_{ij}$. As done previously, taking the weak integral form of $\dot{x} - f(x, u) = 0$

yields [44]

$$\int_{-1}^{1} \phi_i^N(t) \left(\frac{dx^N(t)}{dt} - f(x^N(t), u^N(t)) \right) dt = 0,$$

for $i = 0, 1, \ldots, N$. Integration by parts on the first term results in Galerkin weak form,

$$-\int_{-1}^{1} \frac{d\phi_i^N}{dt} x^N dt + \left[\phi_i^N x^N \right]_{-1}^{1} - \int_{-1}^{1} \phi_i^N f(x^N, u^N) dt = 0.$$

In terms of our approximating polynomials and the true boundary conditions (letting $x^N(-1) \to x(-1)$ and $x^N(1) \to x(1)$) we have

$$-\sum_{j=0}^{N} \int_{-1}^{1} \frac{d\phi_i^N}{dt} \phi_j^N dt\, \bar{x}^{Nj} - \phi_i^N(-1) x(-1) + \phi_i^N(1) x(1) - \int_{-1}^{1} \phi_i^N f(x^N, u^N) dt = 0,$$

for $i = 0, 1, \ldots, N$. Integration by parts, yet again, results in Galerkin strong form with weak enforcement of BCs,

$$\begin{aligned} \sum_{j=0}^{N} \int_{-1}^{1} \phi_i^N \frac{d\phi_j^N}{dt} dt\, \bar{x}^{Nj} + \phi_i^N(-1) \left(\sum_{j=0}^{N} \phi_j^N(-1) \bar{x}^{Nj} - x(-1) \right) \\ -\phi_i^N(1) \left(\sum_{j=0}^{N} \phi_j^N(1) \bar{x}^{Nj} - x(1) \right) - \int_{-1}^{1} \phi_i^N f(x^N, u^N) dt = 0. \end{aligned}$$

The expression may be simplified as

$$\sum_{j=0}^{N} D_{ij} \bar{x}^{Nj} + \kappa_i - \bar{c}^{Ni} = 0,$$

for each $i = 0, 1, \ldots, N$, where the Galerkin differentiation matrix, D, and the RHS vector approximation, $\bar{c}$, are unchanged from those given in the GOCM-S methodology—given by Equations (4.3) and (4.5), respectively—and

$$\kappa_i = \begin{cases} \bar{x}^{N0} - x(-1), & i = 0, \\ x(1) - \bar{x}^{NN}, & i = N, \\ 0, & i \neq 0, N. \end{cases}$$

The BC term κ now provides a natural way to introduce end conditions into our numerical scheme. BCs such as $e(x(-1), x(1)) = \left[x(-1) - x^0, x(1) - x^f\right]^T = [0, 0]^T$ can be imposed by defining κ as

$$\kappa_i = \begin{cases} \bar{x}^{N0} - x^0, & i = 0, \\ x^f - \bar{x}^{NN}, & i = N, \\ 0, & i \neq 0, N, \end{cases}$$

for $i = 0, 1, \ldots, N$.

Remark 5.1. *A similar technique is discussed by Ross et al. in* [11, 84, 85]. *The ability to weakly enforce boundary conditions is not limited to the Galerkin formulation. The collocation method (in the context of the Legendre PS method) may be formulated for weak enforcement of end conditions by modifying the discrete differential Equation 3.5 with a boundary condition term $\tilde{\kappa}$*

$$\sum_{j=0}^{N} A_{ij} \bar{x}^{Nj} + \tilde{\kappa}_i - f(\bar{x}^{Ni}, \bar{u}^{Ni}) = 0, \quad i = 0, 1, \ldots, N, \tag{5.1}$$

where $\tilde{\kappa}$ is given by

$$\tilde{\kappa}_i = \begin{cases} \frac{\bar{x}^{N0} - x^0}{w_i}, & i = 0, \\ \frac{x^f - \bar{x}^{NN}}{w_i}, & i = N, \\ 0, & i \neq 0, N, \end{cases}$$

for fixed end conditions $e(x(-1), x(1)) = \left[x(-1) - x^0, x(1) - x^f\right]^T = [0, 0]^T$.

Remark 5.2. *For existing direct methods for optimal control it is common to enforce the problem's BCs in a strong sense (or exactly) by making them a set of constraints enforced by the nonlinear program (NLP)* [5]*. With the Galerkin optimal control formulation with weak boundary enforcement, BCs can now be enforced in a weak sense. In other words, BCs can be imposed only up to the order of accuracy of the numerical approximation itself, which is sufficient for many applications. In the case of a problem with an incomplete set of end conditions, such as an initial value problem with condition* $e(x(-1), x(1)) = x(-1) - x^0 = 0$*,* κ *may be defined as*

$$\kappa_i = \begin{cases} \bar{x}^{N0} - x^0, & i = 0, \\ 0, & i = N, \\ 0, & i \neq 0, N. \end{cases}$$

Lastly, for more complex BCs such as periodic conditions $e(x(-1), x(1)) = x(-1) - x(1) = 0$*, or other complicated BCs such as nonlinear functions of* $x(-1)$ *and* $x(1)$*,* κ *may be defined as,* $\kappa_i = 0$*, for* $i = 0, 1, \ldots, N$*, and the condition* $e(x(-1), x(1)) = 0$ *may be enforced as a set of constraints by the NLP. However, this last case will result in strong enforcement of the BCs and some of the advantages of using the weak boundary formulation may be lost.*

The dynamical constraint becomes

$$\left\| \sum_{j=0}^{N} D_{ij} \bar{x}^{Nj} + \kappa_i - \bar{c}^{Ni} \right\|_\infty \leq \delta^N, \quad i = 0, 1, \ldots, N.$$

The path constraints are approximated by

$$h(\bar{x}^{Ni}, \bar{u}^{Ni}) \leq \delta^N \cdot \mathbf{1}, \quad i = 0, 1, \ldots, N.$$

Lastly, the cost functional $J[x(\cdot), u(\cdot)]$ is approximated by the LGL quadrature rule,

$$J[x(\cdot), u(\cdot)] \approx \bar{J}^N(\bar{x}^N, \bar{u}^N) = \sum_{i=0}^{N} F(\bar{x}^{Ni}, \bar{u}^{Ni}) w_k + E(\bar{x}^{N0}, \bar{x}^{NN}),$$

where $\bar{x}^N = \left[\bar{x}^{N0},\ \bar{x}^{N1}, \ldots, \bar{x}^{NN}\right]$ and $\bar{u}^N = \left[\bar{u}^{N0},\ \bar{u}^{N1}, \ldots, \bar{u}^{NN}\right]$. To allow for a practical search area for the optimal solution the following constraints are added

$$\{\bar{x}^{Ni} \in \boldsymbol{X}, \bar{u}^{Ni} \in \boldsymbol{U},\ i = 0, 1, \ldots, N\},$$

where $\boldsymbol{X}$ and $\boldsymbol{U}$ are the search regions that contain the optimal solution of the discretized nonlinear optimization.

5.1.2. Computation Strategy

In order to guarantee feasibility of the discretization, Theorems 5.1 and 5.2 show that a relaxation of the dynamical equality constraint to inequality is required, for GOCM-W̃ and GOCM-W, respectively.

5.1.2.1. Computation Strategy for GOCM-W̃

The computational strategy of the GOCM-W̃ is to find the feasible solution $x^N(t) \in \boldsymbol{X}$ and $u^N(t) \in \boldsymbol{U}$ for the following cases:

Case 1. $u(\cdot)$ is piecewise C^0 and $x(\cdot) \in C^0$ and piecewise C^1,

Case 2. $u(\cdot),\ \dot{x}(\cdot) \in H^{m-1}$ and $m \geq 2$,

that minimizes

$$J(x^N(\cdot), u^N(\cdot)) = \int_{-1}^{1} F(x^N(t), u^N(t))dt + E(x^N(-1), x^N(1)), \tag{5.2}$$

subject to the Galerkin constraints

$$\begin{gathered}
\left\| \int_{-1}^{1} \phi_i^N(t) \left(\dot{x}^N(t) - f(x^N(t), u^N(t)) \right) dt + \kappa_i \right\|_\infty \leq MN^{-\alpha}, \quad i = 0, 1, \ldots, N, \\
\left\| h^+(x^N(t), u^N(t)) \right\|_{L^2} \leq MN^{-\alpha},
\end{gathered} \tag{5.3}$$

where $\alpha = \frac{1}{2}$ and $(m-1)$, for Case 1 and 2, respectively; M is a constant independent of N;

$$h^+ = \begin{cases} h, & h > 0, \\ 0, & h \leq 0; \end{cases} \quad \text{and} \quad \kappa_i = \begin{cases} x^N(-1) - x^0, & i = 0, \\ x^f - x^N(1), & i = N, \\ 0, & i \neq 0, N, \end{cases} \tag{5.4}$$

where $e(x(-1), x(1)) = \left[x(-1) - x^0, x(1) - x^f \right]^T = [0, 0]^T$.

5.1.2.2. Computation Strategy for GOCM-W

The computational strategy of the GOCM-W is to find the feasible solution $\bar{x}^N \in \boldsymbol{X}$ and $\bar{u}^N \in \boldsymbol{U}$ for the following cases:

Case 1. $u(\cdot)$ is piecewise C^0 and $x(\cdot) \in C^0$ and piecewise C^1,

Case 2. $u(\cdot),\ \dot{x}(\cdot) \in H^{m-1}$, $m \geq 2$ and $\dot{x}^{(m-1)}(t)$is of bounded variation in $t \in [-1, 1]$,

that minimizes

$$\bar{J}^N(\bar{x}^N, \bar{u}^N) = \sum_{i=0}^{N} F(\bar{x}^{Ni}, \bar{u}^{Ni}) w_i + E(\bar{x}^{N0}, \bar{x}^{NN}), \tag{5.5}$$

subject to the Galerkin constraints

$$\begin{gathered}
\left\| \sum_{j=0}^{N} D_{ij} \bar{x}^{Nj} + \kappa_i - \bar{c}^{Ni} \right\|_\infty \leq MN^{-\alpha}, \quad i = 0, 1, \ldots, N, \\
h(\bar{x}^{Ni}, \bar{u}^{Ni}) \leq MN^{-\alpha} \cdot \mathbf{1}, \quad i = 0, 1, \ldots, N,
\end{gathered} \tag{5.6}$$

where $\alpha = \frac{1}{2}$ and $(m-1)$, for Case 1 and 2, respectively; M is a constant independent of N;

$$\kappa_i = \begin{cases} \bar{x}^{N0} - x^0, & i = 0, \\ x^f - \bar{x}^{NN}, & i = N, \\ 0, & i \neq 0, N, \end{cases} \tag{5.7}$$

where $e(x(-1), x(1)) = \left[x(-1) - x^0, x(1) - x^f\right]^T = [0, 0]^T$.

5.1.3. Feasibility of Solutions

In order to guarantee feasibility of the discretization, Theorems 5.1 and 5.2 introduced in [83] show that a relaxation of the dynamical equality constraint to inequality is required, for GOCM-W̃ and GOCM-W, respectively.

5.1.3.1. Feasibility of GOCM-W̃

Theorem 5.1 (Feasibility of GOCM-W̃)**.** *Given any feasible solution $t \mapsto (x, u)$, for Problem B, consider the following two cases:*

Case 1. $u(\cdot)$ is piecewise C^0 and $x(\cdot) \in C^0$ and piecewise C^1,

Case 2. $u(\cdot)$, $\dot{x}(\cdot) \in H^{m-1}$ and $m \geq 2$.

Then, there exists a positive integer N_0 such that, for any $N \geq N_0$, GOCM-W̃ has a polynomial feasible solution, $(x^N(t), u^N(t))$ such that

$$\left\|x(t) - x^N(t)\right\|_{L^2} \leq MN^{-\alpha} \quad \text{and} \quad \left\|u(t) - u^N(t)\right\|_{L^2} \leq MN^{-\alpha},$$

where $\alpha = \frac{1}{2}$ and $(m-1)$, for Case 1 and 2, respectively; and M is a positive constant independent of N.

Proof. Let $p(t)$ be the $(N-1)$-th order truncated Legendre polynomial approximation of $\dot{x}(t)$. By Lemmas 4.1 and 4.2 there is a constant c_0 independent of N, for any $N \geq N_1$, such that

$$\|\dot{x}(t) - p(t)\|_{L^2} \leq c_0 N^{-\alpha},$$

where $\alpha = \frac{1}{2}$ and $(m-1)$, for Case 1 and 2, respectively. Define

$$x^N(t) = \int_{-1}^{t} p(s)ds + x(-1).$$

Then $p(t) = \dot{x}^N(t)$ and

$$\left\|x(t) - x^N(t)\right\|_{L^2} \leq 2c_0 N^{-\alpha},$$

since, from Hölder's inequality (Lemma 4.3), we have

$$\begin{aligned} \left|x(t) - x^N(t)\right| &= \left|\int_{-1}^{t} (\dot{x}(s) - p(s))\, ds\right| \leq \int_{-1}^{t} |\dot{x}(s) - p(s)| ds \\ \leq &\sqrt{2}\left(\int_{-1}^{1} |\dot{x}(s) - p(s)|^2 ds\right)^{\frac{1}{2}} = \sqrt{2}\|\dot{x}(t) - p(t)\|_{L^2} \leq \sqrt{2} c_0 N^{-\alpha}. \end{aligned}$$

Let $u^N(t)$ be the N-th order Legendre polynomial so that

$$\left\|u(t) - u^N(t)\right\|_{L^2} \leq c_1 N^{-\alpha}.$$

Recall that the LGL quadrature weights, $\{w_k\}_{k=0}^N$, have the property [46],

$$w_k \leq c_2 N^{-1}(1 - (t_k)^2)^{\frac{1}{2}}, \quad k = 1, 2, \ldots, N-1,$$

for constant $c_2 > 0$ independent of k and N; for $k = 0, N$,

$$w_0, w_N = \frac{2}{N(N+1)}.$$

From Theorem 4.1 we have, for each $i = 1, \ldots, N-1$,

$$\begin{aligned} &\left| \int_{-1}^{1} \phi_i^N(t) \left(\dot{x}^N(t) - f(x^N(t), u^N(t)) \right) dt + \kappa_i \right| \\ &= \left| \int_{-1}^{1} \phi_i^N(t) \left(\dot{x}^N(t) - f(x^N(t), u^N(t)) \right) dt \right| \leq c_3 w_i^{\frac{1}{2}} N^{-\alpha}, \end{aligned}$$

for all $N > N_2$, where c_3 is a constant independent of N. For $i = 0$, we have

$$\begin{aligned} &\left| \int_{-1}^{1} \phi_0^N(t) \left(\dot{x}^N(t) - f(x^N(t), u^N(t)) \right) dt + \kappa_0 \right| \\ &\leq \left| \int_{-1}^{1} \phi_0^N(t) \left(\dot{x}^N(t) - f(x^N(t), u^N(t)) \right) dt \right| + \left| x^N(-1) - x(-1) \right| \\ &\leq \left| \int_{-1}^{1} \phi_0^N(t) \left(\dot{x}^N(t) - f(x^N(t), u^N(t)) \right) dt \right| \leq c_3 w_0^{\frac{1}{2}} N^{-\alpha}, \end{aligned}$$

since $\left| x^N(-1) - x(-1) \right| = 0$. For $i = N$, we have

$$\begin{aligned} &\left| \int_{-1}^{1} \phi_N^N(t) \left(\dot{x}^N(t) - f(x^N(t), u^N(t)) \right) dt + \kappa_N \right| \\ &\leq \left| \int_{-1}^{1} \phi_N^N(t) \left(\dot{x}^N(t) - f(x^N(t), u^N(t)) \right) dt \right| + \left| x^N(1) - x(1) \right| \\ &\leq c_3 w_N^{\frac{1}{2}} N^{-\alpha} + \sqrt{2} c_0 N^{-\alpha}, \end{aligned}$$

since

$$\left| x^N(-1) - x(-1) \right| \leq \sqrt{2} \| \dot{x}(t) - p(t) \|_{L^2} \leq \sqrt{2} c_0 N^{-\alpha}.$$

Finally, for each $i = 0, 1, \ldots, N$,

$$\left| \int_{-1}^{1} \phi_k^N(t) \left(\dot{x}^N(t) - f(x^N(t), u^N(t)) \right) dt + \kappa_i \right| \leq M N^{-\alpha},$$

for all $N > N_0$, where M is a constant independent of N.

The estimates for the path constraint follow from the proof of Theorem 4.1.

Thus a solution $(x^N(t), u^N(t))$ to GOCM-W̃ is feasible! □

5.1.3.2. Feasibility of GOCM-W

Theorem 5.2 (Feasibility of GOCM-W). *Given any feasible solution $t \mapsto (x, u)$, for Problem B, consider the following two cases:*

Case 1. $u(\cdot)$ is piecewise C^0 and $x(\cdot) \in C^0$ and piecewise C^1,

Case 2. $u(\cdot),\ \dot{x}(\cdot) \in H^{m-1}$ and $m \geq 2$.

Then, there exists a positive integer N_0 such that, for any $N \geq N_0$, GOCM-W has a feasible solution, $(\bar{x}^N, \bar{u}^N)$ such that

$$\left\|x(t) - x^N(t)\right\|_{L^2} \leq MN^{-\alpha},$$

where $\alpha = \frac{1}{2}$ and $(m-1)$, for Case 1 and 2, respectively; and M is a positive constant independent of N. Additionally, $u^N(t_i) = u(t_i)$, for $i = 0, 1, \ldots, N$.

Proof. Let $p(t)$be the $(N-1)$-th order truncated Legendre polynomial approximation of $\dot{x}(t)$ in the L^2-norm. By Lemma 4.1 there is a constant d_0 independent of N, for any $N \geq N_1$, such that

$$\|\dot{x}(t) - p(t)\|_{L^\infty} \leq d_0 N^{-\beta},$$

where $\beta = \left(m - \frac{3}{2}\right)$, for Case 2. For Case 1 we refer to [77–79] which show the truncated Legendre approximation for discontinuous functions with jump discontinuity (such as the step function defined in Lemma 4.2) displays Gibbs phenomenon. However the maximum amplitude of the overshoot has a finite limit; we conclude that for Case 1, $\beta = 0$. Also, by Lemma 4.2 there is a constant d_1 independent of N, for any $N \geq N_2$, such that

$$\|\dot{x}(t) - p(t)\|_{L^2} \leq d_1 N^{-\alpha},$$

where $\alpha = \frac{1}{2}$ and $(m-1)$, for Case 1 and 2, respectively. Define

$$x^N(t) = \int_{-1}^{t} p(s)ds + x(-1).$$

Then $p(t) = \dot{x}^N(t)$ and

$$\left\| x(t) - x^N(t) \right\|_{L^2} \leq 2d_1 N^{-\alpha},$$

since, from Hölder's inequality (Lemma 4.3),

$$\begin{aligned} \left| x(t) - x^N(t) \right| = & \left| \int_{-1}^{t} \left(\dot{x}(s) - p(s) \right) ds \right| \leq \int_{-1}^{t} |\dot{x}(s) - p(s)| ds \\ \leq & \sqrt{2} \left(\int_{-1}^{1} |\dot{x}(s) - p(s)|^2 ds \right)^{\frac{1}{2}} = \sqrt{2} \|\dot{x}(t) - p(t)\|_{L^2} \leq \sqrt{2} d_1 N^{-\alpha}. \end{aligned}$$

Also, let $u^N(t)$ be an interpolating function of $u(t)$,

$$u^N(t) = \sum_{j=0}^{N} \psi_j^N(t) \bar{u}^{Nj},$$

where $\{\psi_j^N\}_{j=0}^N$ is any set of continuous functions with the property $\psi_j^N(t_i) = \delta_{ij}$, and therefore

$$\bar{u}^{Nj} = u(t_j).$$

Since $x^N(t)$ is a N-th order polynomial, we have

$$\dot{x}^N(t_i) = \sum_{j=0}^{N} A_{ij} \bar{x}^{Nj},$$

where A is the $(N+1) \times (N+1)$ Legendre PS differentiation matrix (2.57) and

$$\bar{x}^{Ni} = x^N(t_i).$$

Recall that the LGL quadrature weights, $\{w_i\}_{i=0}^N$, have the property [46],

$$w_i \leq d_2 N^{-1}(1-(t_i)^2)^{\frac{1}{2}}, \quad i = 1, 2, \ldots, N-1,$$

for constant $d_2 > 0$ independent of i and N; for $i = 0, N$,

$$w_0, w_N = \frac{2}{N(N+1)}.$$

Following from Theorem 4.2, for each $i = 1, 2, \ldots, N-1$,

$$\left| \sum_{j=0}^{N} D_{ij}\bar{x}^{Nj} + \kappa_i - \bar{c}^{Ni} \right| = \left| \sum_{j=0}^{N} D_{ij}\bar{x}^{Nj} - \bar{c}^{Ni} \right| \leq d_3 w_i N^{-\beta},$$

for all $N > N_3$, where d_3 is a constant independent of N. For $i = 0$,

$$\begin{aligned} &\left| \sum_{j=0}^{N} D_{0j}\bar{x}^{Nj} + \kappa_0 - \bar{c}^{N0} \right| \\ \leq &\left| \sum_{j=0}^{N} D_{0j}\bar{x}^{Nj} - \bar{c}^{N0} \right| + \left| x^N(-1) - x(-1) \right| \\ &\leq d_3 w_0 N^{-\beta}, \end{aligned}$$

since $\left|x^N(-1) - x(-1)\right| = 0$. For $i = N$,

$$\begin{aligned}
&\left|\sum_{j=0}^{N} D_{Nj}\bar{x}^{Nj} + \kappa_N - \bar{c}^{NN}\right| \\
&\leq \left|\sum_{j=0}^{N} D_{Nj}\bar{x}^{Nj} - \bar{c}^{NN}\right| + \left|x(1) - x^N(1)\right| \\
&= d_3 w_N N^{-\beta} + \sqrt{2} d_2 N^{-\alpha}.
\end{aligned}$$

Finally, for each $i = 0, 1, \ldots, N$,

$$\left|\sum_{j=0}^{N} D_{ij}\bar{x}^{Nj} + \kappa_i - \bar{c}^{Ni}\right| \leq M N^{-\alpha},$$

for all $N > N_0$, where M is a constant independent of N.

The estimates for the path constraint follow from the proof of Theorem 4.2.

Thus a solution $(\bar{x}^N, \bar{u}^N)$ to GOCM-W is feasible! □

5.2. Galerkin Optimal Control with Continuous Element-based Galerkin

We now consider a continuous element-based Galerkin approach.

5.2.1. Method for Approximation

In this approximation to Problem B, the weak integral form of $\dot{x} - f(x, u) = 0$ in each element, Ω_e, takes the form [44]

$$\int_{\Omega_e} \phi_i^{(e)N}(t) \left(\frac{dx^{(e)N}(t)}{dt} - f(x^{(e)N}(t), u^{(e)N}(t)) \right) dt = 0,$$

where $\Omega = \bigcup_{e=1}^{N_e} \Omega_e$ defines the total domain. The state trajectory, $x(t)$, is approximated inside each element, Ω_e, by interpolating N-th order Lagrange polynomials, $\{\phi_j^{(e)N}(t)\}_{j=0}^{N}$, at the nodes $\{t_j^{(e)}\}_{j=0}^N$ by the relationship

$$x^{(e)N}(t) = \sum_{j=0}^{N} \phi_j^{(e)N}(t) \bar{x}^{(e)Nj},$$

for $e = 1, 2, \ldots, N_e$, where $\{t_j^{(e)}\}_{j=0}^N$ are the LGL nodes, $\{\xi_j\}_{j=0}^N$, mapped back to the physical space inside each element, Ω_e. Also, let $u^N(t)$ be an interpolating function of $\{\bar{u}^{Nj}\}_{j=0}^N$,

$$u^{(e)N}(t) = \sum_{j=0}^{N} \psi_j^{(e)N}(t) \bar{u}^{(e)Nj},$$

where $\{\psi_j^{(e)N}(t)\}_{j=0}^N$ are any set of continuous functions with the property $\psi_j^{(e)N}(t_i) = \delta_{ij}$. Therefore $\bar{x}^{(e)Nj} = x^{(e)N}(t_j^{(e)})$, for $e = 1, 2, \ldots, N_e$ and $j = 0, 1, \ldots, N$, and similarly, $\bar{u}^{(e)Nj} = u^{(e)N}(t_j^{(e)})$. The relationship between the physical time domain, $t \in [t_0, t_f] = \left[t_0^{(1)}, t_N^{(N_e)}\right]$, and the computational space, $\xi \in [-1, 1]$, is given by [44]

$$\xi = \frac{2}{\Delta t^{(e)}} \left(t - t_0^{(e)}\right) - 1 \quad \text{and} \quad d\xi = \frac{2}{\Delta t^{(e)}} dt,$$

and conversely,

$$t = \frac{\Delta t^{(e)}}{2} (\xi + 1) + t_0^{(e)} \quad \text{and} \quad dt = \frac{\Delta t^{(e)}}{2} d\xi,$$

where $\Delta t^{(e)} = t_N^{(e)} - t_0^{(e)}$ is the size of each element, Ω_e, which can be nonuniform in length. The Lagrange polynomial defined on the LGL computational domain is given by

$$\phi_i^N(\xi) = \prod_{\substack{j=0 \\ j \neq i}}^{N} \frac{(\xi - \xi_j)}{(\xi_i - \xi_j)}, \quad i = 0, \ldots, N.$$

The state trajectory, x, can now be approximated inside each element, Ω_e, by

$$x^{(e)N}(\xi) = \sum_{j=0}^{N} \phi_j^N(\xi) \bar{x}^{(e)Nj},$$

where $\{\phi_j^N(\xi)\}_{j=0}^N$ are the Lagrange polynomials defined on the LGL grid. Likewise, $u^N(\xi)$ is given by

$$u^{(e)N}(\xi) = \sum_{j=0}^{N} \psi_j^N(\xi) \bar{u}^{(e)Nj},$$

where $\psi_j^N(\xi_i) = \delta_{ij}$.

Remark 5.3. *In this formulation* $\bar{x}^{(e)NN} = \bar{x}^{(e+1)N0}$ *and* $\bar{u}^{(e)NN} = \bar{u}^{(e+1)N0}$*, for* $e = 1, 2, \ldots, N_e - 1$*. This continuity condition is a consequence of the global formulation of the problem discussed in Remark 5.4.*

In the computational domain, ξ, the system becomes

$$\int_{-1}^{1} \phi_i^N(\xi) \frac{dx^{(e)N}(\xi)}{d\xi} d\xi - \frac{\Delta t^{(e)}}{2} \int_{-1}^{1} \phi_i^N(\xi) f(x^{(e)N}(\xi), u^{(e)N}(\xi)) d\xi = 0,$$

for $e = 1, 2, \ldots, N_e$ and $i = 0, 1, \ldots, N$, and in terms of the approximating polynomials becomes

$$\sum_{j=0}^{N} \int_{-1}^{1} \phi_i^N \frac{d\phi_j^N}{d\xi} d\xi \, \bar{x}^{(e)Nj} - \frac{\Delta t^{(e)}}{2} \int_{-1}^{1} \phi_i^N f(x^{(e)N}, u^{(e)N}) d\xi = 0.$$

In matrix-vector notation, our system can be expressed as

$$\sum_{j=0}^{N} D_{ij}^{(e)} \bar{x}^{(e)Nj} - c_i^{(e)} = 0, \quad i = 0, 1, \ldots, N,$$

for $e = 1, 2, \ldots, N_e$ and $i = 0, 1, \ldots, N$, where the local element $(N+1) \times (N+1)$ Galerkin differentiation matrix, $D^{(e)}$, is the same as that defined in Equation (4.3). If $Q = N$ LGL quadrature nodes are used, the approximation to the $(N+1)\times 1$ RHS vector simplifies to

$$c_i^{(e)} \approx \bar{c}^{(e)Ni} = \frac{\Delta t^{(e)}}{2} f(\bar{x}^{(e)Ni}, \bar{u}^{(e)Ni}) w_i,\ i = 0, 1, \ldots, N,$$

for $e = 1, 2, \ldots, N_e$ and $i = 0, 1, \ldots, N$, where the size of each element, $\Delta t^{(e)}$, can be nonuniform in length.

Remark 5.4. *So far the required objects have been identified to solve the system numerically with element-based Galerkin. However, since nodal basis functions are continuous across element boundaries and LGL nodes include both endpoints, a global solution to our problem can be found. To do this, a global assembly or direct stiffness summation can be done, where the direct stiffness summation operator is* $\bigwedge_{e=1}^{N_e}$. [44]

The global equations to the problem become

$$\sum_{J=1}^{N_p} D_{IJ} \bar{x}^{N_p J} - \bar{c}^{N_p I} = 0, \quad I = 1, \ldots, N_p.$$

The global Galerkin differentiation matrix, D_{IJ} and RHS vector, $\bar{c}^{N_p I}$ are then defined by

$$D_{IJ} = \bigwedge_{e=1}^{N_e} D_{ij}^{(e)}, \quad \text{and} \quad \bar{c}^{N_p I} = \bigwedge_{e=1}^{N_e} \bar{c}^{(e)Ni},$$

where $N_p = (N_e N + 1)$ is the total number of grid points. Note that the direct stiffness summation operator does the mapping $(i, e), (j, e) \to I, J$ [44]. See Section 2.3.2.1 for additional details. The dynamical constraint becomes

$$\left\| \sum_{J=1}^{N_p} D_{IJ}\, \bar{x}^{N_p J} - \bar{c}^{N_p I} \right\|_\infty \leq \delta, \quad I = 1, 2, \ldots, N_p.$$

The endpoint conditions and path constraints are approximated by

$$\begin{aligned}\left\|e(\bar{x}^{N_p0},\bar{x}^{N_pN_p})\right\|_\infty &\le \delta\\ h(\bar{x}^{N_pI},\bar{u}^{N_pI}) &\le \delta\cdot\mathbf{1},\quad I=1,2,\ldots,N_p.\end{aligned}$$

Lastly, the cost functional $J[x(\cdot),u(\cdot)]$ is approximated by the LGL quadrature rule,

$$\begin{aligned}J[x(\cdot),u(\cdot)]\approx&\bar{J}^N(\bar{x}^{N_p},\bar{u}^{N_p})\\ =&\sum_{e=1}^{N_e}\frac{\Delta t^{(e)}}{2}\sum_{i=0}^{N}F(\bar{x}^{N_p((e-1)N+1+i)},\bar{u}^{N_p((e-1)N+1+i)})w_i+E(\bar{x}^{N_p0},\bar{x}^{N_pN_p}),\end{aligned}$$

where $\bar{x}^{N_p}=\left[\bar{x}^{N_p1},\bar{x}^{N_p2},\ldots,\bar{x}^{N_pN_p}\right]$ and $\bar{u}^{N_p}=\left[\bar{u}^{N_p1},\bar{u}^{N_p2},\ldots,\bar{u}^{N_pN_p}\right]$. To allow for a practical search area for the optimal solution the following constraints are included: $\bar{x}^{N_p}\in\boldsymbol{X}$ and $\bar{u}^{N_p}\in\boldsymbol{U}$, where $\boldsymbol{X}$ and $\boldsymbol{U}$ are the search regions that contain the optimal solution of the discretized nonlinear optimization.

5.2.2. Computation Strategy

The computational strategy of the GOCM-CG is to find the feasible solution $\bar{x}^{N_p}\in\boldsymbol{X}$ and $\bar{u}^{N_p}\in\boldsymbol{U}$ that minimizes

$$\begin{aligned}\bar{J}^N(\bar{x}^{N_p},\bar{u}^{N_p})=&\sum_{e=1}^{N_e}\frac{\Delta t^{(e)}}{2}\sum_{i=0}^{N}F(\bar{x}^{N_p((e-1)N+1+i)},\bar{u}^{N_p((e-1)N+1+i)})w_i\\ &+E(\bar{x}^{N_p0},\bar{x}^{N_pN_p}),\end{aligned}$$

subject to the Galerkin constraints

$$\begin{aligned}\left\|\sum_{j=0}^{N_p}D_{IJ}\bar{x}^{N_pJ}-\bar{c}^{NI}\right\|_\infty &\le \delta^N,\quad I=1,2,\ldots,N_p,\\ \left\|e(\bar{x}^{N_p0},\bar{x}^{N_pN_p})\right\|_\infty &\le \delta^N,\\ h(\bar{x}^{N_pI},\bar{u}^{N_pI}) &\le \delta^N\cdot\mathbf{1},\quad I=1,2,\ldots,N_p,\end{aligned}$$

where δ^N is the feasibility tolerance, which is dependent upon N.

5.3. GOCM with Discontinuous Element-based Galerkin

We now consider a discontinuous element-based Galerkin approach introduced in [74].

5.3.1. Method for Approximation

In this approximation to Problem B, the weak integral form of $\dot{x} - f(x,u) = 0$ in each element, Ω_e, yields [44]

$$\int_{\Omega_e} \phi_i^{(e)N}(t) \left(\frac{dx^{(e)N}(t)}{dt} - f(x^{(e)N}(t), u^{(e)N}(t)) \right) dt = 0,$$

where $\Omega = \bigcup_{e=1}^{N_e} \Omega_e$ defines the total domain. The state trajectory, $x(t)$, is approximated inside each element, Ω_e, by interpolating N-th order Lagrange polynomials, $\{\phi_j^{(e)N}(t)\}_{j=0}^N$, at the nodes $\{t_j^{(e)}\}_{j=0}^N$ by the relationship

$$x^{(e)N}(t) = \sum_{j=0}^{N} \phi_j^{(e)N}(t) \bar{x}^{(e)Nj},$$

for $e = 1, 2, \ldots, N_e$, where $\{t_j^{(e)}\}_{j=0}^N$ are the LGL nodes, $\{\xi_j\}_{j=0}^N$, mapped back to the physical space inside each element, Ω_e. Also, let $u^N(t)$ be an interpolating function of $\{\bar{u}^{Nj}\}_{j=0}^N$,

$$u^{(e)N}(t) = \sum_{j=0}^{N} \psi_j^{(e)N}(t) \bar{u}^{(e)Nj},$$

where $\{\psi_j^{(e)N}(t)\}_{j=0}^N$ are any set of continuous functions (not necessarily polynomials) with the property $\psi_j^{(e)N}(t_i) = \delta_{ij}$. Therefore $\bar{x}^{(e)Nj} = x^{(e)N}(t_j^{(e)})$, for $e = 1, 2, \ldots, N_e$ and $j = 0, 1, \ldots, N$, and similarly, $\bar{u}^{(e)Nj} = u^{(e)N}(t_j^{(e)})$. The relationship between the physical time domain, $t \in [t_0, t_f] = \left[t_0^{(1)}, t_N^{(N_e)}\right]$, and the computational space, $\xi \in [-1, 1]$, is given by [44]

$$\xi = \frac{2}{\Delta t^{(e)}}\left(t - t_0^{(e)}\right) - 1 \quad \text{and} \quad d\xi = \frac{2}{\Delta t^{(e)}} dt,$$

and conversely,

$$t = \frac{\Delta t^{(e)}}{2}(\xi + 1) + t_0^{(e)} \quad \text{and} \quad dt = \frac{\Delta t^{(e)}}{2} d\xi,$$

where $\Delta t^{(e)} = t_N^{(e)} - t_0^{(e)}$ is the size of each element, Ω_e, which can be nonuniform in length. The Lagrange polynomial defined on the LGL computational domain is given by

$$\phi_i^N(\xi) = \prod_{\substack{j=0 \\ j \neq i}}^{N} \frac{(\xi - \xi_j)}{(\xi_i - \xi_j)}, \quad i = 0, \ldots, N.$$

The state trajectory, x, can now be approximated inside each element, Ω_e, by

$$x^{(e)N}(\xi) = \sum_{j=0}^{N} \phi_j^N(\xi) \bar{x}^{(e)Nj},$$

where $\{\phi_j^N(\xi)\}_{j=0}^N$ are the Lagrange polynomials defined on the LGL grid. Likewise, $u^N(\xi)$ is given by

$$u^{(e)N}(\xi) = \sum_{j=0}^{N} \psi_j^N(\xi) \bar{u}^{(e)Nj},$$

where $\psi_j^N(\xi_i) = \delta_{ij}$. In the computational domain, ξ, the system becomes

$$\int_{-1}^{1} \phi_i^N(\xi) \frac{dx^{(e)N}(\xi)}{d\xi} d\xi - \frac{\Delta t^{(e)}}{2} \int_{-1}^{1} \phi_i^N(\xi) f(x^{(e)N}(\xi), u^{(e)N}(\xi)) d\xi = 0,$$

for $e = 1, 2, \ldots, N_e$ and $i = 0, 1, \ldots, N$. Integration by parts on the first term yields the weak form relationship

$$-\int_{-1}^{1} \frac{d\phi_i^N}{d\xi} x^{(e)N} d\xi + \left[\phi_i^N x^{(e)N}\right]_{-1}^{1} - \frac{\Delta t^{(e)}}{2} \int_{-1}^{1} \phi_i f(x^{(e)N}, u^{(e)N}) d\xi = 0,$$

and in terms of our approximating polynomials, we have

$$-\sum_{j=0}^{N} \int_{-1}^{1} \frac{d\phi_i^N}{d\xi} \phi_j^N d\xi \, \bar{x}^{(e)Nj} + \sum_{j=0}^{N} \left[\phi_i^N \phi_j^N\right]_{1}^{-1} \bar{x}_j^{(*)} - \frac{\Delta t^{(e)}}{2} \int_{-1}^{1} \phi_i^N f(x^{(e)N}, u^{(e)N}) d\xi = 0.$$

Remark 5.5. *With the discontinuous element-based Galerkin approach, we let $\dot{x}$, u and the basis functions be discontinuous across element edges. A numerical flux term $\bar{x}^{(*)}$ acts as a jump condition between elements* [44]. *Here, we consider the centered flux relationship, $\bar{x}^{(*)} = \frac{1}{2}\left(\bar{x}^{(e)} + \bar{x}^{(q)}\right)$, proposed by Delfour et al.* [66], *where e and q denote the element and its neighbor, respectively.*

Integrating by parts, yet again, results in the Galerkin strong form relationship

$$\sum_{j=0}^{N} \int_{-1}^{1} \phi_i^N \frac{d\phi_j^N}{d\xi} d\xi \, \bar{x}^{(e)Nj} + \eta_i^{(e)} - \frac{\Delta t^{(e)}}{2} \int_{-1}^{1} \phi_i^N f(x^{(e)N}, u^{(e)N}) d\xi = 0,$$

for $e = 1, 2, \ldots, N_e$ and $i = 0, 1, \ldots, N$. Since LGL nodes are used, the boundary term, $\eta^{(e)}$, may be simplified as

$$\eta_i^{(1)} = \begin{cases} \frac{1}{2}\left(\bar{x}^{(2)N0} - \bar{x}^{(1)NN}\right), & i = N, \\ 0, & i \neq N, \end{cases}$$

$$\eta_i^{(N_e)} = \begin{cases} \frac{1}{2}\left(\bar{x}^{(N_e)N0} - \bar{x}^{(N_e-1)NN}\right), & i = 0, \\ 0, & i \neq 0, \end{cases}$$

for elements $\Omega_e = \Omega_1$ and Ω_{N_e}, respectively, and for each other element $(\Omega_e \neq \Omega_1, \Omega_{N_e})$ we have

$$\eta_i^{(e)} = \begin{cases} \frac{1}{2}\left(\bar{x}^{(e)N0} - \bar{x}^{(e-1)NN}\right), & i = 0, \\ \frac{1}{2}\left(\bar{x}^{(e+1)N0} - \bar{x}^{(e)NN}\right), & i = N, \\ 0, & i \neq 0, N. \end{cases}$$

Remark 5.6. *The problem's endpoint conditions have not been introduced into the boundary term, $\eta^{(e)}$, and will instead be enforce in a strong sense through a set of endpoint constraints, as done in GOCM-S. If instead BCs are to be enforced weakly, a modification can be made to the boundary condition term, $\eta^{(e)}$.*

In matrix-vector notation, our system may be expressed as

$$\sum_{j=0}^{N} D_{ij}^{(e)} \bar{x}_j^{(e)} + \eta_i^{(e)} - c_i^{(e)} = 0, \quad e = 1, 2, \ldots, N_e,\ i = 0, 1, \ldots, N,$$

where the local element $(N+1) \times (N+1)$ Galerkin differentiation matrix, $D^{(e)}$, is the same as that defined in Equation (4.3). If $Q = N$ LGL quadrature nodes are used, the approximation to the $(N+1) \times 1$ RHS vector simplifies to

$$c_i^{(e)} \approx \bar{c}^{(e)Ni} = \frac{\Delta t^{(e)}}{2} f(\bar{x}^{(e)Ni}, \bar{u}^{(e)Ni}) w_i,\ e = 1, 2, \ldots, N_e,\ i = 0, 1, \ldots, N,$$

where the size of each element, $\Delta t^{(e)}$, can be nonuniform in length. The dynamical constraint becomes

$$\left\| \sum_{j=0}^{N} D_{ij}^{(e)} \bar{x}^{(e)Nj} - \eta_i^{(e)} - \bar{c}^{(e)Ni} \right\|_{\infty} \leq \delta, \quad e = 1, 2, \ldots, N_e,\ i = 0, 1, \ldots, N.$$

The endpoint conditions and path constraints are approximated by

$$\begin{aligned} &\left\| e(\bar{x}_0^{(1)}, \bar{x}_N^{(N_e)}) \right\|_\infty \leq \delta, \\ &\qquad h(\bar{x}_i^{(e)}, \bar{u}_i^{(e)}) \leq \delta \cdot \mathbf{1}, \quad e = 1, 2, \ldots, N_e,\ i = 0, 1, \ldots, N, \end{aligned}$$

Lastly, the cost functional $J[x(\cdot), u(\cdot)]$ is approximated by LGL quadrature rule,

$$J[x(\cdot), u(\cdot)] \approx \bar{J}^N(\bar{x}^N, \bar{u}^N) = \sum_{e=1}^{N_e} \frac{\Delta t^{(e)}}{2} \sum_{k=0}^{N} F(\bar{x}^{(e)Nk}, \bar{u}^{(e)Nk}) w_k + E(\bar{x}^{(1)N0}, \bar{x}^{(N_e)NN}),$$

where

$$\begin{aligned} \bar{x}^N &= \left[\bar{x}^{(1)N0}, \ldots, \bar{x}^{(1)NN}, \ldots, \bar{x}^{(N_e)N0}, \ldots, \bar{x}^{(N_e)NN}\right], \\ \bar{u}^N &= \left[\bar{u}^{(1)N0}, \ldots, \bar{u}^{(1)NN}, \ldots, \bar{u}^{(N_e)N0}, \ldots, \bar{u}^{(N_e)NN}\right]. \end{aligned}$$

To allow for a practical search area for the optimal solution the following constraints are included: $\bar{x}^N \in \boldsymbol{X}$ and $\bar{u}^N \in \boldsymbol{U}$, where $\boldsymbol{X}$ and $\boldsymbol{U}$ are the search regions that contain the optimal solution of the discretized nonlinear optimization.

5.3.2. Computation Strategy

The computational strategy of the GOCM-DG is to find the feasible solution $\bar{x}^N \in \boldsymbol{X}$ and $\bar{u}^N \in \boldsymbol{U}$ that minimizes

$$\bar{J}^N(\bar{x}^N, \bar{u}^N) = \sum_{e=1}^{N_e} \frac{\Delta t^{(e)}}{2} \sum_{i=0}^{N} F(\bar{x}^{(e)Ni}, \bar{u}^{(e)Ni}) w_i + E(\bar{x}^{(1)N0}, \bar{x}^{(N_e)NN}),$$

subject to the Galerkin constraints

$$\begin{aligned}\left\|\sum_{j=0}^{N} D_{ij}^{(e)}\bar{x}^{(e)Nj} - \eta_i^{(e)} - \bar{c}^{(e)Ni}\right\|_\infty &\leq \delta^N, \quad e = 1, 2, \ldots, N_e,\ i = 0, 1, \ldots, N,\\ \left\|e(\bar{x}^{(1)N0}, \bar{x}^{(N_e)NN})\right\|_\infty &\leq \delta^N,\\ h(\bar{x}^{(e)Ni}, \bar{u}^{(e)Ni}) &\leq \delta^N \cdot \mathbf{1}, \quad e = 1, 2, \ldots, N_e,\ i = 0, 1, \ldots, N,\end{aligned}$$

where δ^N is the feasibility tolerance, which is dependent upon N.

5.4. Galerkin Optimal Control for Multi-scale Problems

In this section, a multi-scale Galerkin optimal control approach is proposed to solve a specific optimal control problem, one in which the system dynamics are of different timescales (see Problem $\tilde{\text{B}}$). Such a problem may consists of a fast state, $x_f(t)$, associated with the fast dynamics and a slow state, $x_s(t)$, associated with the slow dynamics. GOCM-S (see Chapter 4) serves as the basis for the Galerkin multi-scale approach to Problem $\tilde{\text{B}}$. In particular, system (4.6) is fundamental to this alternative Galerkin optimal control formulation described below. The outline shown here follows the strategy given by Gong et al. in [71].

5.4.1. Method for Approximation

The states and controls are approximated with globally interpolating Lagrange polynomials on different LGL timescales. The slow state, $x_s(t)$, is approximated on sparse grid $\{\tau_j\}_{j=0}^{M}$ while the fast state, $x_f(t)$, on dense grid $\{t_j\}_{j=0}^{N}$, where $M < N$. The slow and fast states are defined by the following approximating polynomials

$$x_s(t) \approx x_s^M(t) = \sum_{j=0}^{M} \phi_j^M(t)\bar{x}_s^{Mj},$$

$$x_f(t) \approx x_f^N(t) = \sum_{j=0}^{N} \phi_j^N(t)\bar{x}_f^{Nj},$$

where the Lagrange polynomials $\{\phi_j^M(t)\}_{j=0}^M$ and $\{\phi_j^N(t)\}_{j=0}^N$ are defined on grids $\{\tau_j\}_{j=0}^M$ and $\{t_j\}_{j=0}^N$, respectively. Let

$$\begin{aligned}\bar{x}_s^{Mj} &\approx x_s(\tau_j), \quad j = 0, 1, \ldots, M,\\ \bar{x}_f^{Nj} &\approx x_f(t_j), \quad j = 0, 1, \ldots, N,\end{aligned}$$

and similarly, $\bar{u}^{Nj} \approx u(t_j)$, for $j = 0, 1, \ldots, N$.

Remark 5.7. *For simplicity, the control variable, $u(t)$, is approximated on the dense grid $\{t_i\}_{i=0}^N$, however this need not be the case. Modifications may be made to the process outlined in this section to cast the control onto a unique grid, such as sparse grid $\{\tilde{\tau}_k\}_{k=0}^{\tilde{M}}$, where $\tilde{M} < N$* [71].

For GOCM-MS, a solution to the differential equations $\dot{x}_s = f(x_s, x_f, u)$ and $\dot{x}_f = g(x_s, x_f, u)$ may be approximated by discretizing the slow dynamics over the dense grid with the following formulation

$$\begin{aligned}\sum_{j=0}^{M} D_{ij}^{NM} \bar{x}_s^{Mj} - \bar{c}_s^{Ni} &= 0, \quad i = 0, 1, \ldots, N,\\ \sum_{j=0}^{N} D_{ij}^{NN} \bar{x}_f^{Nj} - \bar{c}_f^{Ni} &= 0, \quad i = 0, 1, \ldots, N.\end{aligned}$$

The $(N+1) \times (M+1)$, *non-square*, differentiation transformation matrix D^{NM} and $(N+1) \times (N+1)$, *square*, differentiation matrix D^{NN} are defined by

$$\begin{aligned}D_{ij}^{NM} &= \int_{-1}^{1} \phi_i^N \frac{d\phi_j^M}{dt} dt = \frac{d\phi_j^M}{dt}(t_i) w_i = A_{ij}^{NM} w_i, \quad i = 0, 1, \ldots, N,\ j = 0, 1, \ldots, M,\\ D_{ij}^{NN} &= \int_{-1}^{1} \phi_i^N \frac{d\phi_j^N}{dt} dt = \frac{d\phi_j^N}{dt}(t_i) w_i = A_{ij}^{NN} w_i, \quad i, j = 0, 1, \ldots, N,\end{aligned}$$

where $\{w_i\}_{i=0}^N$ are the LGL weights associated with LGL points $\{t_i\}_{i=0}^N$, A^{NN} is the standard $(N+1) \times (N+1)$ Legendre PS differentiation matrix (2.57) and A^{NM} is the $(N+1) \times (M+1)$ Legendre PS differentiation transformation matrix (2.39). The RHS vectors, $\bar{c}_s^N$ and $\bar{c}_f^N$, are defined by

$$\begin{aligned}\bar{c}_s^{Ni} &= f(\hat{x}_s^{Ni}, \bar{x}_f^{Ni}, \bar{u}^{Ni})w_i = 0, \quad i = 0, 1, \dots, N,\\ \bar{c}_f^{Ni} &= g(\hat{x}_s^{Ni}, \bar{x}_f^{Ni}, \bar{u}^{Ni})w_i = 0, \quad i = 0, 1, \dots, N.\end{aligned}$$

The slow state approximation projected to the dense grid, $\hat{x}_s^N$, may be calculated by the linear mapping $T_{ij}^{NM} = \phi_j^M(t_i)$ with the relationship

$$\hat{x}_s^{Ni} = \sum_{j=0}^{n} T_{ij}^{NM} \bar{x}_s^{Mj},$$

for $i = 0, 1, \dots, N$, where T^{NM} is the $(N+1) \times (M+1)$ transformation matrix (2.37).

Remark 5.8. *Projecting the slow dynamics onto the dense grid provides a way of capturing the high frequency information of the fast state. If instead the intuitive approach is used, of discretizing the slow dynamics over the sparse grid, the high frequency information of the fast state is lost, resulting in a decrease in method accuracy* [71].

The dynamical constraints therefore become

$$\begin{aligned}\left\| \sum_{j=0}^{M} D_{ij}^{NM} \bar{x}_s^{Mj} - \bar{c}_s^{Ni} \right\|_\infty &\leq \delta^N, \quad i = 0, 1, \dots, N,\\ \left\| \sum_{j=0}^{N} D_{ij}^{NN} \bar{x}_f^{Nj} - \bar{c}_f^{Ni} \right\|_\infty &\leq \delta^N, \quad i = 0, 1, \dots, N.\end{aligned}$$

The endpoint conditions and path constraints are approximated similarly by

$$\begin{aligned}\left\| e(\bar{x}_s^{M0}, \bar{x}_s^{MM}, \bar{x}_f^{N0}, \bar{x}_f^{NN}) \right\|_\infty &\leq \delta^N,\\ h(\hat{x}_s^{Ni}, \bar{x}_f^{Ni}, \bar{u}^{Ni}) &\leq \delta^N \cdot \mathbf{1}, \quad i = 0, 1, \dots, N.\end{aligned}$$

Lastly, the cost functional $J[x(\cdot), u(\cdot)]$ is approximated by the LGL quadrature rule,

$$J[x(\cdot), u(\cdot)] \approx \bar{J}^N(\bar{x}_s^M, \bar{x}_f^N, \bar{u}^N) = \sum_{i=0}^{N} F(\hat{x}_s^{Ni}, \bar{x}_f^{Ni}, \bar{u}^{Ni}) w_i + E(\bar{x}_s^{M0}, \bar{x}_s^{MM}, \bar{x}_f^{N0}, \bar{x}_f^{NN}),$$

where

$$\bar{x}_s^M = \left[\bar{x}_s^{M0},\ \bar{x}_s^{M1}, \ldots, \bar{x}_s^{MM}\right], \quad \bar{x}_f^N = \left[\bar{x}_f^{N0},\ \bar{x}_f^{N1}, \ldots, \bar{x}_f^{NN}\right]$$
$$\text{and} \quad \bar{u}^N = \left[\bar{u}^{N0},\ \bar{u}^{N1}, \ldots, \bar{u}^{NN}\right].$$

To allow for a search area for the optimal solution the following constraints are included: $\bar{x}_s^M \in \boldsymbol{X}_s$, $\bar{x}_f^N \in \boldsymbol{X}_f$ and $\bar{u}^N \in \boldsymbol{U}$, where $\boldsymbol{X}_s$, $\boldsymbol{X}_f$ and $\boldsymbol{U}$ are the search regions that contain the optimal solution of the discretized nonlinear optimization.

5.4.2. Computation Strategy

The computational strategy of the GOCM-MS is to find the feasible solution $\bar{x}_s^M \in \boldsymbol{X}_s$, $\bar{x}_f^N \in \boldsymbol{X}_f$ and $\bar{u}^N \in \boldsymbol{U}$ that minimizes

$$\bar{J}^N(\bar{x}_s^M, \bar{x}_f^N, \bar{u}^N) = \sum_{i=0}^{N} F(\hat{x}_s^{Ni}, \bar{x}_f^{Ni}, \bar{u}^{Ni}) w_i + E(\bar{x}_s^{M0}, \bar{x}_s^{MM}, \bar{x}_f^{N0}, \bar{x}_f^{NN}),$$

subject to the Galerkin constraints

$$\begin{aligned}
\left\| \sum_{j=0}^{M} D_{ij}^{NM} \bar{x}_s^{Mj} - \bar{c}_s^{Ni} \right\|_\infty &\leq \delta^N, \quad i = 0, 1, \ldots, N, \\
\left\| \sum_{j=0}^{N} D_{ij}^{NN} \bar{x}_f^{Nj} - \bar{c}_f^{Ni} \right\|_\infty &\leq \delta^N, \quad i = 0, 1, \ldots, N, \\
\left\| e(\bar{x}_s^{M0}, \bar{x}_s^{MM}, \bar{x}_f^{N0}, \bar{x}_f^{NN}) \right\|_\infty &\leq \delta^N, \\
h(\hat{x}_s^{Ni}, \bar{x}_f^{Ni}, \bar{u}^{Ni}) &\leq \delta^N \cdot \mathbf{1}, \quad i = 0, 1, \ldots, N,
\end{aligned}$$

where δ^N is the feasibility tolerance, which is dependent upon N.

5.5. Modifications to Galerkin Optimal Control

Two modifications to the general Galerkin optimal control formulation will be discussed in this section: over-integration of the RHS vector and the use of quadrature points other than LGL, such as LG and LGR points.

5.5.1. Over-Integration of the RHS Vector

Thus far, all the Galerkin optimal control formulations discussed have approximated the RHS vector with inexact integration, and $N+1$ quadrature points. It may, however, be advantageous to approximate the RHS vector with increased accuracy. This may be accomplished by over-integration of the RHS vector while approximating the integral using LGL quadrature. As outlined in the GOCM-S, the state trajectory, $x(t)$, is approximated by

$$x(t) \approx x^N(t) = \sum_{j=0}^{N} \phi_j^N(t)\bar{x}^{Nj}.$$

Let

$$\bar{x}^{Nj} \approx x(t_j), \quad j = 0, 1, \ldots, N,$$

and similarly, $\bar{u}^{Nj}$ is the approximation of $u(t_j)$. In the Galerkin optimal control formulation, a solution to the differential equation $\dot{x} - f(x,u) = 0$ may be approximated at the LGL nodes with the following weak integral formulation [44]

$$\int_{-1}^{1} \phi_i^N(t) \left(\frac{dx^N(t)}{dt} - f(x^N(t), u^N(t)) \right) dt = 0,$$

and in terms of the approximating polynomials becomes

$$\sum_{j=0}^{N} \int_{-1}^{1} \phi_i^N \frac{d\phi_j^N}{dt} dt\, \bar{x}^{Nj} - \int_{-1}^{1} \phi_i^N f(x^N, u^N) dt = 0.$$

In matrix-vector form the system becomes

$$\sum_{j=0}^{N} D_{ij}\bar{x}^{Nj} - c_i = 0, \quad i = 0, 1, \ldots, N,$$

where the $(N+1) \times (N+1)$ differentiation matrix D is defined by Equation (4.2) and calculated exactly by Equation (4.3). The $(N+1) \times 1$ RHS vector c is defined as

$$c_i = \int_{-1}^{1} \phi_i^N(t) f(x^N(t), u^N(t))dt, \quad i = 0, 1, \ldots, N,$$

and can be approximated with LGL quadrature by the relationship

$$c_i \approx \bar{c}^{Ni} = \sum_{k=0}^{Q} \phi_i^N(t_k) f(x^N(t_k), u^N(t_k)) w_k, \quad i = 0, 1, \ldots, N,$$

where $\{w_k\}_{k=0}^{N}$ are the LGL quadrature weights (2.49) associated with LGL points, $\{t_k\}_{k=0}^{N}$.

Remark 5.9. *Recall that for LGL quadrature rule, $Q = N + 1$ integration points will integrate the RHS vector exactly when $f(x(t), u(t))$ is linear in $x(t)$ and $u(t)$. In the case of a nonlinear function, f, the accuracy of the numerical integration of the RHS vector may also be improved if $Q = N + 1$ integration points are used when f consists of one or more nonlinear terms.*

Using over-integration, the RHS vector approximation becomes

$$\bar{c}^{Ni} = \sum_{k=0}^{N+1} \phi_i^N(t_k) f(x^N(t_k), u^N(t_k)) w_k, \quad i = 0, 1, \ldots, N,$$

where $\{w_k\}_{k=0}^{N+1}$ are the LGL quadrature weights associated with LGL points, $\{t_k\}_{k=0}^{N+1}$.The dynamical constraint therefore becomes

$$\left\| \sum_{j=0}^{N} D_{ij}\bar{x}^{Nj} - \bar{c}^{Ni} \right\|_{\infty} \leq \delta^N, \quad i = 0, 1, \ldots, N.$$

The endpoint conditions and path constraints are approximated similarly by

$$\begin{aligned} \left\| e(\bar{x}^{N0}, \bar{x}^{NN}) \right\|_\infty &\leq \delta^N, \\ h(\bar{x}^{Ni}, \bar{u}^{Ni}) &\leq \delta^N \cdot \mathbf{1}, \quad i = 0, 1, \ldots, N, \end{aligned}$$

where $\mathbf{1}$ denotes $[1, \ldots, 1]^T$. Lastly, the cost functional $J[x(\cdot), u(\cdot)]$ is approximated by the LGL quadrature rule,

$$J[x(\cdot), u(\cdot)] \approx \bar{J}^N(\bar{x}^N, \bar{u}^N) = \sum_{i=0}^{N} F(\bar{x}^{Ni}, \bar{u}^{Ni}) w_i + E(\bar{x}^{N0}, \bar{x}^{NN}),$$

where $\bar{x}^N = \left[\bar{x}^{N0}\, \bar{x}^{N1} \cdots \bar{x}^{NN}\right]$ and $\bar{u}^N = \left[\bar{u}^{N0}\, \bar{u}^{N1} \cdots \bar{u}^{NN}\right]$. To allow for a practical search area for the optimal solution the following constraints are included: $\bar{x}^N \in \boldsymbol{X}$ and $\bar{u}^N \in \boldsymbol{U}$.

5.5.2. Galerkin Optimal Control with LG and LGR/F-LGR Quadrature Points

The Galerkin weak formulation with weak boundary condition enforcement also allows for consideration of quadrature points that do not include both endpoints, e.g., LG and LGR (or F-LGR) nodes. This is advantageous since LG and LGR quadrature rules may lead to increased accuracies when performing the RHS vector integration. Recall that LG quadrature rule integration is exact for polynomial integrands of degree less than or equal to $2N + 1$, where the LG nodes, $\{t_k\}_{k=0}^{N}$, are defined by $-1 < t_0 < \cdots < t_N < 1$ and are the roots of Equation (2.41). LGR (and F-LGR) quadrature rule is exact for polynomial integrands of degree less than or equal to 2N, where the LGR nodes, $\{t_k\}_{k=0}^{N}$, are defined by $t_0 = -1 < t_1 < \cdots < t_N < 1$, and are the roots of Equation (2.43) and the F-LGR nodes are the negative of the LGR points.

5.5.2.1. Method for Approximation for Galerkin Optimal Control with LGR/F-LGR Nodes

In this section, a formulation is proposed to solve a specific optimal control problem, one in which the initial or final conditions of the problem dynamics are provided (not both). Consider Problem B such that one of the two following cases exist:

Case 1: $e(x(-1), x(1)) = x(-1) - x^0 = 0$,

Case 2: $e(x(-1), x(1)) = x(1) - x^f = 0$,

where x^0 and x^f are constants. As with the GOCM-W, we will consider the weak enforcement of the end conditions, however, now we will use the Galerkin weak form. In this approximation to Problem B, the state trajectory, $x(t)$, is approximated with globally interpolating N-th order Lagrange polynomials, $\{\phi_j^N\}_{j=0}^N$, defined on a grid of F-LGR or LGR nodes, $\{t_j\}_{j=0}^N$, for Case 1 and 2, respectively,

$$x(t) \approx x^N(t) = \sum_{j=0}^{N} \phi_j^N(t) \bar{x}^{Nj}.$$

Due to the property of the Lagrange polynomials, $\phi_j^N(t_i) = \delta_{ij}$, we have

$$\bar{x}^{Nj} = x^N(t_j), \quad j = 0, 1, \dots, N.$$

Also, let $u^N(t)$ be an interpolating function of $\{\bar{u}^{Nj}\}_{j=0}^N$,

$$u^N(t) = \sum_{j=0}^{N} \psi_j^N(t) \bar{u}^{Nj},$$

where $\{\psi_j^N\}_{j=0}^N$ is any set of continuous functions with the property $\psi_j^N(t_i) = \delta_{ij}$. Taking the weak integral form of $\dot{x} - f(x, u) = 0$ yields [44]

$$\int_{-1}^{1} \phi_i^N(t) \left(\frac{dx^N(t)}{dt} - f(x^N(t), u^N(t)) \right) dt = 0,$$

for $i = 0, 1, \ldots, N$. Integration by parts on the first term results in Galerkin weak form,

$$-\int_{-1}^{1} \frac{d\phi_i^N}{dt} x^N dt + \left[\phi_i^N x^N\right]_{-1}^{1} - \int_{-1}^{1} \phi_i^N f(x^N, u^N) dt = 0.$$

In terms of our approximating polynomials, we have

$$-\sum_{j=0}^{N} \int_{-1}^{1} \frac{d\phi_i^N}{dt} \phi_j^N dt\, \bar{x}^{Nj} - \phi_i^N(-1) x^N(-1) + \phi_i^N(1) x^N(1) - \int_{-1}^{1} \phi_i^N f(x^N, u^N) dt = 0,$$

for $i = 0, 1, \ldots, N$. The expression may be simplified as

$$\sum_{j=0}^{N} \tilde{D}_{ij} \bar{x}^{Nj} + \tilde{\kappa}_i - \bar{c}^{Ni} = 0,$$

for each $i = 0, 1, \ldots, N$, where

$$\tilde{\kappa}_i = \begin{cases} -\phi_i^N(-1) x^0, & i \neq N, \\ -\phi_N^N(-1) x^0 + \bar{x}^{NN}, & i = N, \end{cases}$$

for Case 1 and

$$\tilde{\kappa}_i = \begin{cases} \phi_0^N(1) x^f - \bar{x}^{N0}, & i = 0, \\ \phi_i^N(1) x^f, & i \neq 0, \end{cases}$$

for Case 2.

The $(N+1) \times (N+1)$ differentiation matrix $\tilde{D}$ is defined by

$$\tilde{D}_{ij} = -\int_{-1}^{1} \frac{d\phi_i^N(t)}{dt} \phi_j^N(t) dt, \quad i, j = 0, 1, \ldots, N.$$

If F-LGR/LGR quadrature rule is used with $Q = N$ quadrature points, the differentiation matrix, $\tilde{D}$, can be calculated exactly by the relationship

$$\tilde{D}_{ij} = -\sum_{k=0}^{Q} \frac{d\phi_i^N(t_k)}{dt}\phi_j^N(t_k) w_k = -\frac{d\phi_i^N}{dt}(t_j) w_j = -A_{ij}^T w_j, \quad i, j = 0, 1, \ldots, N,$$

where $\{w_j\}_{j=0}^N$ are the F-LGR/LGR quadrature weights (2.50) and $A_{ij} = \dot{\phi}_j^N(t_i)$ is the F-LGR/LGR PS differentiation matrix.

The RHS vector, c can be approximated by the relationship

$$c_i \approx \bar{c}^{Ni} = \sum_{k=0}^{Q} \phi_i^N(t_k) f(x^N(t_k), u^N(t_k)) w_k = f(\bar{x}^{Ni}, \bar{u}^{Ni}) w_i, \quad i = 0, 1, \ldots, N.$$

Remark 5.10. *Recall that for F-LGR/LGR quadrature rule, integration is exact for polynomial integrands of degree less than or equal to $2N$. If $Q = N$ integration points are used, the RHS vector will integrate exactly when $f(x(t), u(t))$ is linear in $x(t)$ and $u(t)$. In the case of a nonlinear function f, the accuracy of integration may be improved by increasing the number of quadrature points Q (See Section 2.2.2).*

The dynamical constraint becomes

$$\left\| \sum_{j=0}^{N} \tilde{D}_{ij} \bar{x}^{Nj} + \tilde{\kappa}_i - \bar{c}^{Ni} \right\|_{\infty} \leq \delta^N, \quad i = 0, 1, \ldots, N.$$

The path constraints are approximated by

$$h(\bar{x}^{Ni}, \bar{u}^{Ni}) \leq \delta^N \cdot \mathbf{1}, \quad i = 0, 1, \ldots, N.$$

Lastly, the cost functional $J[x(\cdot), u(\cdot)]$ is approximated by the F-LGR/LGR quadrature rule,

$$J[x(\cdot), u(\cdot)] \approx \bar{J}^N(\bar{x}^N, \bar{u}^N) = \sum_{i=0}^{N} F(\bar{x}^{Ni}, \bar{u}^{Ni}) w_i + E(\bar{x}^{N0}, \bar{x}^{NN}),$$

where $\bar{x}^N = \left[\bar{x}^{N0}\, \bar{x}^{N1} \cdots \bar{x}^{NN}\right]$ and $\bar{u}^N = \left[\bar{u}^{N0}\, \bar{u}^{N1} \cdots \bar{u}^{NN}\right]$. To allow for a practical search area for the optimal solution the following constraints are included: $\bar{x}^N \in \boldsymbol{X}$ and $\bar{u}^N \in \boldsymbol{U}$.

Remark 5.11. *Note that the use of the Galerkin optimal control formulation outlined here with LGR or F-LGR nodes does not automatically provide the optimal control solutions at one endpoint of the domain (this applies to $t = 1$ for LGR points and $t = -1$ for F-LGR points). However, if required by the application, the control solutions at $t = 1$ or $t = -1$ may be found by interpolation of the control approximation, $\bar{u}^N$ with a possible reduction in accuracy.*

5.5.2.2. Method for Approximation for Galerkin Optimal Control with LG Nodes

In this section, a formulation is proposed to solve a specific optimal control problem, one in which a complete set of boundary conditions are provided for the problem dynamics. Consider Problem B such that

$$e(x(-1), x(1)) = \left[x(-1) - x^0, x(1) - x^f\right]^T = [0, 0]^T,$$

where x^0 and x^f are constants. As with the GOCM-W, we will consider the weak enforcement of the boundary conditions, however, now we will use the Galerkin weak form. In this approximation to Problem B, the state trajectory, $x(t)$, is approximated with globally interpolating N-th order Lagrange polynomials,$\{\phi_j^N\}_{j=0}^N$, defined on a grid of LG nodes, $\{t_j\}_{j=0}^N$,

$$x(t) \approx x^N(t) = \sum_{j=0}^{N} \phi_j^N(t)\bar{x}^{Nj}.$$

Due to the property of the Lagrange polynomials, $\phi_j^N(t_i) = \delta_{ij}$, we have

$$\bar{x}^{Nj} = x^N(t_j), \quad j = 0, 1, \ldots, N.$$

Also, let $u^N(t)$ be an interpolating function of $\{\bar{u}^{Nj}\}_{j=0}^N$,

$$u^N(t) = \sum_{j=0}^{N} \psi_j^N(t)\bar{u}^{Nj},$$

where $\{\psi_j^N\}_{j=0}^N$ is any set of continuous functions (not necessarily polynomials) with the property $\psi_j^N(t_i) = \delta_{ij}$. Taking the weak integral form of $\dot{x} - f(x,u) = 0$ yields [44]

$$\int_{-1}^{1} \phi_i^N(t) \left(\frac{dx^N(t)}{dt} - f(x^N(t), u^N(t)) \right) dt = 0,$$

for $i = 0, 1, \ldots, N$. Integration by parts on the first term results in Galerkin weak form,

$$-\int_{-1}^{1} \frac{d\phi_i^N}{dt} x^N dt + \left[\phi_i^N x^N\right]_{-1}^{1} - \int_{-1}^{1} \phi_i^N f(x^N, u^N) dt = 0.$$

In terms of our approximating polynomials (and introducing the true initial condition, $x^N(-1) \to x(-1)$ and $x^N(1) \to x(1)$) we have

$$-\sum_{j=0}^{N} \int_{-1}^{1} \frac{d\phi_i^N}{dt} \phi_j^N dt\, \bar{x}^{Nj} - \phi_i^N(-1)x(-1) + \phi_i^N(1)x(1) - \int_{-1}^{1} \phi_i^N f(x^N, u^N) dt = 0,$$

for $i = 0, 1, \ldots, N$. By letting $x(-1) = x^0$ and $x(1) = x^f$, the expression may be simplified as

$$\sum_{j=0}^{N} \tilde{D}_{ij} \bar{x}^{Nj} + \tilde{\kappa}_i - \bar{c}^{Ni} = 0,$$

for each $i = 0, 1, \ldots, N$, where

$$\tilde{\kappa}_i = -\phi_i^N(-1)x^0 + \phi_i^N(1)x^f.$$

The $(N+1)\times(N+1)$ differentiation matrix, $\tilde{D}$, is defined by

$$\tilde{D}_{ij} = -\int_{-1}^{1} \frac{d\phi_i^N(t)}{dt}\phi_j^N(t)dt, \quad i,j = 0,1,\dots,N.$$

If LG quadrature rule is used with $Q = N$ quadrature points, the differentiation matrix, $\tilde{D}$, can be calculated exactly by the relationship

$$\tilde{D}_{ij} = -\sum_{k=0}^{Q} \frac{d\phi_i^N(t_k)}{dt}\phi_j^N(t_k)w_k = -\frac{d\phi_i^N}{dt}(t_j)w_j = -A_{ij}^T w_j, \quad i,j = 0,1,\dots,N,$$

where $\{w_j\}_{j=0}^N$ are the LG quadrature weights (2.48) and $A_{ij} = \dot{\phi}_j^N(t_i)$ is the LG PS differentiation matrix.

The RHS vector, c can be approximated by the relationship

$$c_i \approx \bar{c}^{Ni} = \sum_{k=0}^{Q} \phi_i^N(t_k) f(x^N(t_k), u^N(t_k))w_k = f(\bar{x}^{Ni}, \bar{u}^{Ni})w_i, \quad i = 0,1,\dots,N.$$

Remark 5.12. *Recall that for LG quadrature rule, integration is exact for polynomial integrands of degree less than or equal to $2N+1$. If $Q = N$ integration points are used, the RHS vector will integrate exactly when $f(x(t), u(t))$ is linear in $x(t)$ and $u(t)$. In the case of a nonlinear function f, the accuracy of integration may be improved by increasing the number of quadrature points Q (See Section 2.2.2).*

The dynamical constraint becomes

$$\left\| \sum_{j=0}^{N} \tilde{D}_{ij}\bar{x}^{Nj} + \tilde{\kappa}_i - \bar{c}^{Ni} \right\|_\infty \leq \delta^N, \quad i = 0,1,\dots,N.$$

The path constraints are approximated by

$$h(\bar{x}^{Ni}, \bar{u}^{Ni}) \leq \delta^N \cdot \mathbf{1}, \quad i = 0,1,\dots,N.$$

Lastly, the cost functional $J[x(\cdot), u(\cdot)]$ is approximated by the LG quadrature rule,

$$J[x(\cdot), u(\cdot)] \approx \bar{J}^N(\bar{x}^N, \bar{u}^N) = \sum_{i=0}^{N} F(\bar{x}^{Ni}, \bar{u}^{Ni}) w_i + E(\bar{x}^{N0}, \bar{x}^{NN}),$$

where $\bar{x}^N = \left[\bar{x}^{N0}\, \bar{x}^{N1} \cdots \bar{x}^{NN}\right]$ and $\bar{u}^N = \left[\bar{u}^{N0}\, \bar{u}^{N1} \cdots \bar{u}^{NN}\right]$. To allow for a practical search area for the optimal solution the following constraints are included: $\bar{x}^N \in \boldsymbol{X}$ and $\bar{u}^N \in \boldsymbol{U}$.

Remark 5.13. *Note that the use of the Galerkin optimal control formulation outlined here with LG nodes does not automatically provide the optimal control solutions at the endpoints of the domain, $t = \pm 1$. However, if required by the application, the control solutions at $t = -1$ and/or $t = 1$ may be found by interpolation of the control approximation, $\bar{u}^N$ with a possible reduction in accuracy.*

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 6:
GALERKIN OPTIMAL CONTROL WITH LEGENDRE POLYNOMIAL TEST FUNCTIONS

In this chapter, a Galerkin optimal control formulation is proposed where Legendre polynomials replace Lagrange polynomials as test functions in the weak integral approximation of the problem dynamics. The purpose of this modification is highlighted in consistency Theorem 6.2, which is valid for problems with discontinuous controls (unlike consistency Theorems 4.3 and 4.4 for Problems GOCM-S̃ and GOCM-S, respectively). Theorem 6.2 proves that the nonlinear programming Problems GOCM-$\tilde{L}$ is a consistent approximation to the continuous optimal control Problem B, even those with *piecewise continuous* controls.

6.1. Methods for Approximation

In the Galerkin optimal control approximation to Problem B, with Legendre polynomial test functions, the state trajectory, $x(t)$, is approximated with globally interpolating N-th order Lagrange polynomials, $\{\phi_j^N\}_{j=0}^N$, defined on a grid of LGL nodes, $\{t_j\}_{j=0}^N$,

$$x(t) \approx x^N(t) = \sum_{j=0}^{N} \phi_j^N(t)\bar{x}^{Nj}.$$

Due to the property of the Lagrange polynomials, $\phi_j^N(t_i) = \delta_{ij}$, we have

$$\bar{x}^{Nj} = x^N(t_j), \quad j = 0, 1, \ldots, N.$$

Also, let $u^N(t)$ be an interpolating function of $\{\bar{u}^{Nj}\}_{j=0}^N$,

$$u^N(t) = \sum_{j=0}^{N} \psi_j^N(t)\bar{u}^{Nj},$$

where $\{\psi_j^N\}_{j=0}^N$ is any set of continuous functions (not necessarily polynomials) with the property $\psi_j^N(t_i) = \delta_{ij}$. In this formulation, a solution to the differential equation $\dot{x} - f(x,u) = 0$ may be approximated at the LGL nodes with the following weak integral formulation

$$\int_{-1}^{1} \tilde{L}_i(t) \left(\frac{dx^N(t)}{dt} - f\left(x^N(t), u^N(t)\right) \right) dt = 0,$$

for $i = 0, 1, \ldots, N$, where the test functions, $\tilde{L}_i = \frac{L_i}{\|L_i\|_{L^2}}$, are the normalized Legendre polynomials of order i. The L^2-norm of L_i is given by [72] as

$$\|L_i\|_{L^2} = \sqrt{\frac{2}{2i+1}}.$$

In terms of the approximating polynomials, the system becomes

$$\sum_{j=0}^{N} \int_{-1}^{1} \tilde{L}_i \frac{d\phi_j^N}{dt} dt\, \bar{x}^{Nj} - \int_{-1}^{1} \tilde{L}_i f(x^N, u^N) dt = 0.$$

In matrix-vector form, the system becomes

$$\sum_{j=0}^{N} D_{ij}^L \bar{x}^{Nj} - c_i^L = 0, \quad i = 0, 1, \ldots, N,$$

where the $(N+1) \times (N+1)$ differentiation matrix D^L is defined by

$$D_{ij}^L = \int_{-1}^{1} \tilde{L}_i(t) \frac{d\phi_j^N(t)}{dt} dt, \quad i, j = 0, 1, \ldots, N,$$

and the $(N+1) \times 1$ right-hand-side (RHS) vector c is defined as

$$c_i^L = \int_{-1}^{1} \tilde{L}_i(t) f(x^N(t), u^N(t)) dt, \quad i = 0, 1, \ldots, N.$$

If LGL quadrature rule is used, with $Q = N$ quadrature points, the differentiation matrix, D^L, can be calculated exactly by the relationship

$$D^L_{ij} = \sum_{k=0}^{N} \tilde{L}_i(t_k) \frac{d\phi_j^N}{dt}(t_k) w_k, \quad i, j = 0, 1, \ldots, N.$$

If $Q = N$ quadrature points are used, the RHS vector, c^L, can be approximated by the relationship

$$c^L_i \approx \bar{c}^{Ni}_L = \sum_{k=0}^{N} \tilde{L}_i(t_k) f(\bar{x}^{Nk}, \bar{u}^{Nk}) w_k, \quad i = 0, 1, \ldots, N.$$

The system may be simplified as

$$\sum_{j=0}^{N} D^L_{ij} \bar{x}^{Nj} - \bar{c}^{Ni}_L = 0, \quad i = 0, 1, \ldots, N.$$

The dynamical constraint therefore becomes

$$\left\| \sum_{j=0}^{N} D^L_{ij} \bar{x}^{Nj} - \bar{c}^{Ni}_L \right\|_\infty \leq \delta^N, \quad i = 0, 1, \ldots, N.$$

The endpoint conditions and path constraints are approximated similarly by

$$\begin{aligned} \left\| e(\bar{x}^{N0}, \bar{x}^{NN}) \right\|_\infty &\leq \delta^N, \\ h(\bar{x}^{Ni}, \bar{u}^{Ni}) &\leq \delta^N \cdot \mathbf{1}, \quad i = 0, 1, \ldots, N. \end{aligned}$$

Lastly, the cost functional $J[x(\cdot), u(\cdot)]$ is approximated by the LGL quadrature rule,

$$J[x(\cdot), u(\cdot)] \approx \bar{J}^N(\bar{x}^N, \bar{u}^N) = \sum_{i=0}^{N} F(\bar{x}^{Ni}, \bar{u}^{Ni}) w_i + E(\bar{x}^{N0}, \bar{x}^{NN}),$$

where $\bar{x}^N = \left[\bar{x}^{N0}\,\bar{x}^{N1}\cdots\bar{x}^{NN}\right]$ and $\bar{u}^N = \left[\bar{u}^{N0}\,\bar{u}^{N1}\cdots\bar{u}^{NN}\right]$. To allow for a practical search area for the optimal solution the following constraints are included: $\bar{x}^N \in \boldsymbol{X}$ and $\bar{u}^N \in \boldsymbol{U}$, where $\boldsymbol{X}$ and $\boldsymbol{U}$ are the search regions that contain the optimal solution of the discretized nonlinear optimization.

6.2. Computation Strategy

The computation strategy for Galerkin optimal control with Legendre polynomial test functions is presented in two forms. First, the strategy for the continuous problem, in terms of the approximating polynomials is outlined, denoted as GOCM-$\tilde{L}$. Next, the discrete problem, discretized on a LGL grid is presented, denoted as GOCM-L.

6.2.1. Computation Strategy for GOCM-$\tilde{L}$

The computational strategy of the GOCM-$\tilde{L}$ is to find the feasible solution $x^N(t) \in \boldsymbol{X}$ and $u^N(t) \in \boldsymbol{U}$ for the following cases:

Case 1. $u(\cdot)$ is piecewise C^0 and $x(\cdot) \in C^0$ and piecewise C^1,

Case 2. $u(\cdot),\ \dot{x}(\cdot) \in H^{m-1}$ and $m \geq 2$,

that minimizes

$$J(x^N(\cdot), u(\cdot)) = \int_{-1}^{1} F\left(x^N(t), u(t)\right) dt + E\left(x^N(-1), x^N(1)\right),$$

subject to the Galerkin constraints

$$\begin{aligned} \left\| \int_{-1}^{1} \tilde{L}_i(t)\left(\dot{x}^N(t) - f\left(x^N(t), u(t)\right)\right) dt \right\|_{\infty} &\leq MN^{-\alpha}, \quad i = 0, 1, \ldots, N, \\ \left\| e\left(x^N(-1), x^N(1)\right) \right\|_{\infty} &\leq MN^{-\alpha}, \\ \left\| h^{+}\left(x^N(t), u^N(t)\right) \right\|_{L^2} &\leq MN^{-\alpha}, \end{aligned}$$

where $\alpha = \frac{1}{2}$ and $(m-1)$, for Case 1 and 2, respectively; M is a constant independent of N and

$$h^{+} = \begin{cases} h, & h > 0, \\ 0, & h \leq 0. \end{cases}$$

6.2.2. Computation Strategy for GOCM-L

The computational strategy of the GOCM-L is to find the feasible solution $\bar{x}^N(t) \in \boldsymbol{X}$ and $\bar{u}^N(t) \in \boldsymbol{U}$ that minimizes

$$\bar{J}^N\left(\bar{x}^N, \bar{u}^N\right) = \sum_{i=0}^{N} F\left(\bar{x}^{Ni}, \bar{u}^{Ni}\right) w_i + E\left(\bar{x}^{N0}, \bar{x}^{NN}\right),$$

subject to the Galerkin constraints

$$\begin{aligned} \left\| \sum_{j=0}^{N} D_{ij}^{L} \bar{x}^{Nj} - \bar{c}_L^{Ni} \right\|_{\infty} &\leq \delta^N, \quad i = 0, 1, \ldots, N, \\ \left\| e\left(\bar{x}^{N0}, \bar{x}^{NN}\right) \right\|_{\infty} &\leq \delta^N, \\ h\left(\bar{x}^{Ni}, \bar{u}^{Ni}\right) &\leq \delta^N \cdot \mathbf{1}, \quad i = 0, 1, \ldots, N. \end{aligned}$$

6.3. Feasibility of Solutions

Theorem 6.1 (Feasibility of GOCM-$\tilde{L}$)**.** *Given any feasible solution $t \mapsto (x, u)$, for Problem B, consider the following two cases:*

Case 1. $u(\cdot)$ is piecewise C^0 and $x(\cdot) \in C^0$ and piecewise C^1,

Case 2. $u(\cdot),\ \dot{x}(\cdot) \in H^{m-1}$ and $m \geq 2$.

Then, there exists a positive integer N_0 such that, for any $N \geq N_0$, GOCM-$\tilde{\text{L}}$ has a polynomial feasible solution, $(x^N(t), u^N(t))$ such that

$$\left\| x(t) - x^N(t) \right\|_{L^2} \leq M N^{-\alpha} \quad \text{and} \quad \left\| u(t) - u^N(t) \right\|_{L^2} \leq M N^{-\alpha},$$

where $\alpha = \frac{1}{2}$ and $(m-1)$, for Case 1 and 2, respectively; and M is a positive constant independent of N.

Proof. Let $p(t)$ be the $(N-1)$-th order truncated Legendre polynomial approximation of $\dot{x}(t)$. By Lemmas 4.1 and 4.2 there is a constant c_0 independent of N, for any $N \geq N_0$, such that

$$\|\dot{x}(t) - p(t)\|_{L^2} \leq c_0 N^{-\alpha},$$

where $\alpha = \frac{1}{2}$ and $(m-1)$, for Case 1 and 2, respectively. Define

$$x^N(t) = \int_{-1}^{t} p(s)ds + x(-1).$$

Then $p(t) = \dot{x}^N(t)$ and

$$\left\|x(t) - x^N(t)\right\|_{L^2} \leq 2c_0 N^{-\alpha},$$

since, from Hölder's inequality (Lemma 4.3), we have

$$\begin{aligned} \left|x(t) - x^N(t)\right| &= \left|\int_{-1}^{t} \left(\dot{x}(s) - p(s)\right) ds\right| \leq \int_{-1}^{t} |\dot{x}(s) - p(s)| ds \\ \leq& \sqrt{2} \left(\int_{-1}^{1} |\dot{x}(s) - p(s)|^2 ds\right)^{\frac{1}{2}} = \sqrt{2} \|\dot{x}(t) - p(t)\|_{L^2} \leq \sqrt{2} c_0 N^{-\alpha}. \end{aligned} \quad (6.1)$$

Let $u^N(t)$ be the N-th order Legendre polynomial so that

$$\left\|u(t) - u^N(t)\right\|_{L^2} \leq c_1 N^{-\alpha}.$$

From our Galerkin approximation, Hölder's inequality (Lemma 4.3), and the property, $\|L_i\|_{L^2} = \sqrt{\frac{2}{2i+1}}$, we have for each $i = 0, 1, \ldots, N$,

$$
\begin{aligned}
&\left|\int_{-1}^{1} \tilde{L}_i(t)\left(\dot{x}^N(t) - f\left(x^N(t), u^N(t)\right)\right) dt\right| \\
&\leq \left\|\frac{L_i(t)}{\sqrt{\frac{2}{2i+1}}}\right\|_{L^2} \left\|\dot{x}^N(t) - f\left(x^N(t), u^N(t)\right)\right\|_{L^2} \\
&= \left\|\dot{x}^N(t) - f\left(x^N(t), u^N(t)\right)\right\|_{L^2} \\
&\leq \left\|\dot{x}(t) - \dot{x}^N(t)\right\|_{L^2} + \left\|f\left(x(t), u(t)\right) - f(x^N(t), u^N(t))\right\|_{L^2} \\
&= c_0 N^{-\frac{1}{2}} + l_1\left\|x(t) - x^N(t)\right\|_{L^2} + l_2\left\|u(t) - u^N(t)\right\|_{L^2} \\
&\leq c_0 N^{-\alpha} + 2l_1 c_0 N^{-\alpha} + l_2 c_1 N^{-\alpha},
\end{aligned}
$$

where l_1 and l_2 are the Lipschitz constants of f with respect to x and u, respectively, which are independent of N. It follows that

$$
\left|\int_{-1}^{1} \tilde{L}_i(t)\left(\dot{x}^N(t) - f\left(x^N(t), u^N(t)\right)\right) dt\right| \leq M N^{-\alpha},
$$

and holds for each $i = 0, 1, \ldots, N$, and all $N > N_0$, where M is a constant independent of N.

For the endpoint condition we have

$$
\begin{aligned}
\left|x(1) - x^N(1)\right| &= \left|\int_{-1}^{t} \left(\dot{x}(s) - p(s)\right) ds\right| \leq \int_{-1}^{t} |\dot{x}(s) - p(s)| ds \\
\leq &\sqrt{2}\left(\int_{-1}^{1} |\dot{x}(s) - p(s)|^2 ds\right)^{\frac{1}{2}} = \sqrt{2}\|\dot{x}(t) - p(t)\|_{L^2} \leq \sqrt{2} c_0 N^{-\alpha},
\end{aligned}
$$

so we have, by Lipschitz condition,

$$
\left|e(x^N(-1), x^N(1))\right| \leq M N^{-\alpha}.
$$

For the path constraint let $\mathcal{D} = \left\{t|h\left(x^N(t), u^N(t)\right) > 0\right\}, \overline{\mathcal{D}} = [-1,1] \backslash \mathcal{D}$, since $h\left(x(t), u(t)\right) \le 0$. Then

$$\begin{aligned}\left\|h^+\left(x^N(t), u^N(t)\right)\right\|_{L^2} &= \left(\int_{\mathcal{D}} \left(h(x^N(t), u^N(t))\right)^2 dt\right)^{\frac{1}{2}}\\ &\le \left(\int_{\mathcal{D}} \left(h\left(x^N(t), u^N(t)\right) - h\left(x(t), u(t)\right)\right)^2 dt\right)^{\frac{1}{2}}\\ \le \Bigg(\int_{\mathcal{D}}(h(x^N(t), u^N(t)) - h(x(t), u(t)))^2 dt &+ \int_{\overline{\mathcal{D}}}(h(x^N(t), u^N(t)) - h(x(t), u(t)))^2 dt\Bigg)^{\frac{1}{2}}\\ &= \left(\int_{-1}^{1} \left(h\left(x^N(t), u^N(t)\right) - h\left(x(t), u(t)\right)\right)^2 dt\right)^{\frac{1}{2}}\\ &= \left\|h\left(x^N(t), u^N(t)\right) - h\left(x(t), u(t)\right)\right\|_{L^2}\\ &\le l_3\left\|x(t) - x^N(t)\right\|_{L^2} + l_4\left\|u(t) - u^N(t)\right\|_{L^2} \le MN^{-\alpha},\end{aligned}$$

where l_3 and l_4 are the Lipschitz constants of h with respect to x and u, respectively, which are independent of N. Hence

$$\left\|h^+\left(x^N(t), u^N(t)\right)\right\|_{L^2} \le MN^{-\alpha}.$$

Thus a solution $\left(x^N(t), u^N(t)\right)$ to GOCM-$\tilde{L}$ is feasible! □

6.4. Consistency of Solutions

Theorem 6.2 (Consistency of GOCM-$\tilde{L}$)**.** *Suppose $\left(x^N(t), u^N(t)\right)$ is a solution of GOCM-$\tilde{L}$ and there exists $(x(t), u(t))$ such that:*

Case 1. $u(\cdot)$ is piecewise C^0 and $x(\cdot) \in C^0$ and piecewise C^1,

Case 2. $u(\cdot),\ \dot{x}(\cdot) \in H^{m-1}$ and $m \ge 2$.

Also, suppose

$$\lim_{N\to\infty}\left\|u(t) - u^N(t)\right\|_{L^2} = 0, \qquad (6.2)$$

and $x^N(t) \to x(t)$ uniformly, thus

$$\lim_{N\to\infty} \left\| x(t) - x^N(t) \right\|_{L^2} = 0. \tag{6.3}$$

Then $(x(t), u(t))$ satisfies

$$\begin{cases} \|\dot{x}(t) - f\left(x(t), u(t)\right)\|_{L^2} = 0, \\ e\left(x(-1), x(1)\right) = 0, \\ h\left(x(t), u(t)\right) \leq 0, \end{cases}$$

and is an optimal solution to Problem B.

Proof. Due to the completeness of the Legendre polynomials [80], to prove $\|\dot{x}(t) - f\left(x, u\right)\|_{L^2} = 0$ it is sufficient to prove

$$\int_{-1}^{1} \tilde{L}_i(t)\left(\dot{x}(t) - f\left(x(t), u(t)\right)\right) dt = 0,$$

for each $i = 0, 1, \ldots, \infty$ (see definition of completeness, Definition 4.2). Consider

$$\begin{aligned} &\left| \int_{-1}^{1} \tilde{L}_i(t)\left(\dot{x}(t) - f\left(x(t), u(t)\right)\right) dt \right| \\ \leq &\left| \int_{-1}^{1} \tilde{L}_i(t)\left(\dot{x}^N(t) - f\left(x^N(t), u^N(t)\right)\right) dt \right| + \left| \int_{-1}^{1} \tilde{L}_i(t)\left(\dot{x}(t) - \dot{x}^N(t)\right) dt \right| \\ &+ \left| \int_{-1}^{1} \tilde{L}_i(t)\left(f\left(x^N(t), u^N(t)\right) - f\left(x, u\right)\right) dt \right| \\ \leq &\, MN^{-\alpha} + \left\| \dot{x}(t) - \dot{x}^N(t) \right\|_{L^2} + \left\| f\left(x(t), u(t)\right) - f\left(x^N(t), u(t)\right) \right\|_{L^2} \\ = &\, MN^{-\alpha} + \left\| x(t) - x^N(t) \right\|_{L^2} + l_1 \left\| x(t) - x^N(t) \right\|_{L^2} + l_2 \left\| u(t) - u^N(t) \right\|_{L^2}, \end{aligned}$$

where $\alpha = \frac{1}{2}$ and $(m - 1)$, for Case 1 and 2, respectively (from Theorem 6.1); M is a positive constant and l_1 and l_2 are the Lipschitz constants of f with respect to x and u, respectively, all independent of N. It follows that as $N \to \infty$ we have

$$\|\dot{x}(t) - f\left(x(t), u(t)\right)\|_{L^2} = 0.$$

For the endpoint condition, since $x^N(t) \to x(t)$ uniformly, we have $x^N(1) \to x(1)$ and $x^N(-1) \to x(-1)$. Since, from the formulation of the computational strategy we have $\left|e\left(x^N(-1), x^N(1)\right)\right| \leq MN^{-\alpha}$, we conclude that $e\left(x(-1), x(1)\right) = 0$ as $N \to \infty$.

For the path constraint, since $h(x(t), u(t))$ is piecewise C^1, if $h\left(x(t^*), u(t^*)\right) > 0$, $\exists$ an interval (a, b) in which $h\left(x(t), u(t)\right) > 0$. Then

$$\|h\left(x(t), u(t)\right)\|_{L^2(a,b)} = \left(\int_a^b \left(h\left(x(t), u(t)\right)\right)^2 dt\right)^{\frac{1}{2}} > 0.$$

However,

$$\begin{aligned}\|h\left(x(t), u(t)\right)\|_{L^2(a,b)} &\leq \left\|h\left(x(t), u(t)\right) - h^+\left(x^N(t), u^N(t)\right)\right\|_{L^2(a,b)} + \left\|h^+\left(x^N(t), u^N(t)\right)\right\|_{L^2(a,b)} \\ &\leq \left\|h\left(x(t), u(t)\right) - h\left(x^N(t), u^N(t)\right)\right\|_{L^2(a,b)} + MN^{-\alpha} \\ &\leq l_3\left\|x(t) - x^N(t))\right\|_{L^2} + l_4\left\|u(t) - u^N(t))\right\|_{L^2} + MN^{-\alpha},\end{aligned}$$

where l_3 and l_4 are the Lipschitz constants of h with respect to x and u, respectively, which are independent of N. Hence, this is a contradiction, therefore $h\left(x(t), u(t)\right) \leq 0$ as $N \to \infty$.

Suppose that $(x(t), u(t))$ is not optimal. Then $\exists$ $\left(x^*(t), u^*(t)\right)$ so that

$$J\left(x^*(\cdot), u^*(\cdot)\right) < J\left(x(\cdot), u(\cdot)\right).$$

Also, $\exists$ $\left(x^*(t), u^*(t)\right)$ such that

$$\left\|x^{*N}(t) - x(t)\right\|_{L^2} \leq MN^{-\alpha} \text{ and } \left\|u^{*N}(t) - u(t)\right\|_{L^2} \leq MN^{-\alpha},$$

where $\big(x^{*N}(t), u^{*N}(t)\big)$ is a feasible trajectory of GOCM-$\tilde{L}$. Therefore

$$J\left(x^{*N}(\cdot), u^{*N}(\cdot)\right) \geq J\left(x^{N}(\cdot), u^{N}(\cdot)\right). \tag{6.4}$$

However,

$$\begin{aligned} &\big|J\left(x^{N}(\cdot), u^{N}(\cdot)\right) - J\left(x(\cdot), u(\cdot)\right)\big| \\ \leq &\int_{-1}^{1}\big|F\left(x^{N}(t), u^{N}(t)\right) - F\left(x(t), u(t)\right)\big|dt + \big|E\left(x^{N}(-1), x^{N}(1)\right) - x(-1), x(1))\big| \\ \leq &\sqrt{2}\big\|F\left(x^{N}(t), u^{N}(t)\right) - F\left(x(t), u(t)\right)dt\big\|_{L^2} + \big|E\left(x^{N}(-1), x^{N}(1)\right) - x(-1), x(1))\big|. \end{aligned}$$

Due to the Lipschitz condition and assumptions (6.2) and (6.3) we have

$$\lim_{N\to\infty}\big|J\left(x^{N}(\cdot), u^{N}(\cdot)\right) - J\left(x(\cdot), u(\cdot)\right)\big| = 0.$$

Similarly,

$$\lim_{N\to\infty}\big|J\left(x^{*N}(\cdot), u^{*N}(\cdot)\right) - J\left(x^{*}(\cdot), u^{*}(\cdot)\right)\big| = 0.$$

Therefore, from (6.4) we have

$$J\left(x^{*}(\cdot), u^{*}(\cdot)\right) \geq J\left(x(\cdot), u(\cdot)\right).$$

This is a contradiction, since we assumed

$$J\left(x^{*}(\cdot), u^{*}(\cdot)\right) < J\left(x(\cdot), u(\cdot)\right).$$

We conclude that $(x(t), u(t))$ achieves an optimal cost and therefore is an optimal solution to Problem B! □

Remark 6.1. *Theorems 6.1 and 6.2 show that solutions to the GOCM-$\tilde{L}$ exist and provide confidence that they will converge to the optimal solutions of Problem B. More importantly, Theorem 6.2 provides a foundation to show that solutions to the GOCM-$\tilde{L}$ will converge to optimal solutions with discontinuities in the control, such as solutions to bang-bang control problems. However, as with the GOCM-$\tilde{S}$, questions still remain about the conditions under which the underlying assumptions exist (in the case of the GOCM-$\tilde{L}$, assumptions (6.2) and (6.3)), but the required analysis is above the scope of this dissertation.*

CHAPTER 7:
EXAMPLE PROBLEMS

A number of examples are provided in this chapter to highlight the versatility and accuracy of the Galerkin optimal control formulations discussed in Chapters 4–6. Examples 7.1, 7.2 and 7.4 demonstrate the improved accuracy provided by the Galerkin optimal control formulation with weak enforcement of boundary conditions over strong enforcement for problems with fixed boundary conditions as well as incomplete sets of end conditions. In particular, the examples show that the GOCM-W has an advantage over the GOCM-S for low order approximations of control solutions. Examples 7.1 and 7.4 also show the potential advantages of the Galerkin formulations with F-LGR and LG points, respectively. Additionally, Example 7.3 demonstrates the effectiveness of the element-based Galerkin optimal control formulations (such as the GOCM-DG) when employed to approximated optimal control problems with discontinuous controls. Lastly, Examples 7.5 and 7.6 demonstrate the computational efficiency in which the multi-scale Galerkin optimal control formulation may solve multi-scale problems, those with states and controls that evolve on different timescales. In contrast to the difficulties with the multi-scale Legendre PS method (see Section 3.4) highlighted in Chapter 3, the GOCM-MS is shown to successfully reduce the size of multi-scale problems.

7.1. Example 7.1: Nonlinear Two-Dimensional Problem with Fixed Initial Conditions

Consider the *nonlinear* two-dimensional problem with *fixed initial conditions* given by Gong et al. [3] of minimizing the cost function

$$J = 4x_1(2) + x_2(2) + 4\int_0^2 u^2 dt, \tag{7.1}$$

subject to the dynamics

$$\dot{x}_1(t) = x_2^3(t) \quad \text{and} \quad \dot{x}_2(t) = u(t),$$

and with initial conditions

$$x_1(0) = 0 \quad \text{and} \quad x_2(0) = 1.$$

The analytic solution to this problem is given by

$$\begin{aligned} x_1(t) &= \frac{2}{5} - \frac{64}{5(2+t)^5}, \\ x_2(t) &= \frac{4}{(2+t)^2}, \\ u(t) &= \frac{-8}{(2+t)^3}, \end{aligned}$$

obtained via Pontryagin's maximum principle. This problem was solved using the GOCM-W (Section 5.1) with optimality and feasibility tolerances of 10^{-8} and 10^{-8}, respectively. A two-point initial guess was provided. Figure 16 shows a comparison of the exact solution with the GOCM-W approximation and $N = 20$.

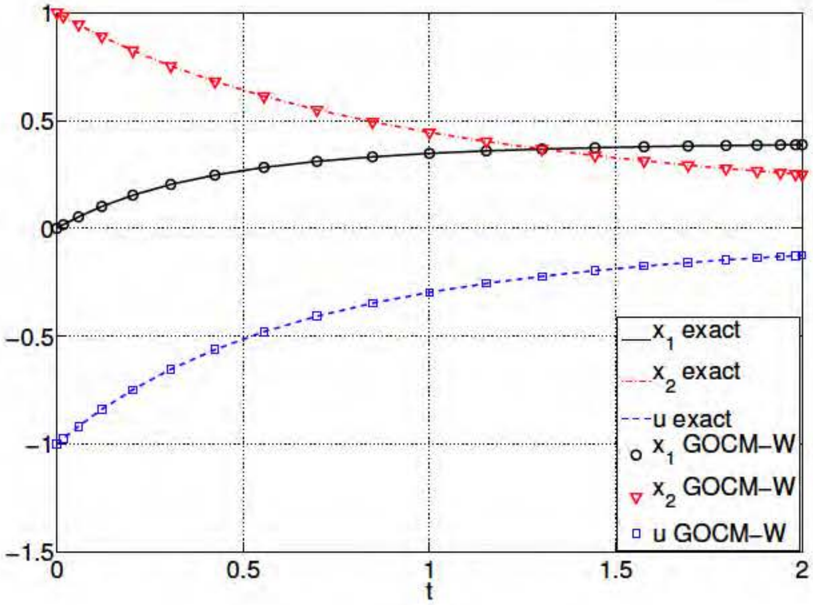


Figure 16: Exact solution and GOCM-W approximation with $N = 20$ for Example 7.1.

Figures 17, 18 and 19 show the maximum error of the states and control,

$$\begin{aligned}
\|\text{error}_{x_1}\|_\infty &= \left\|x_1(t_i) - x_1^N(t_i)\right\|_\infty, \quad i = 0, 1, \ldots, N,\\
\|\text{error}_{x_2}\|_\infty &= \left\|x_2(t_i) - x_1^N(t_i)\right\|_\infty, \quad i = 0, 1, \ldots, N,\\
\|\text{error}_{u}\|_\infty &= \left\|u(t_i) - u^N(t_i)\right\|_\infty, \quad i = 0, 1, \ldots, N,
\end{aligned}$$

respectively, for the LPM (Legendre PS method, Section 3.2), GOCM-S (Chapter 4), GOCM-W and GOCM-FLGR (Section 5.5.2.1) vs. polynomial order, N. Note that the GOCM-S approximations show nearly the exact same exponential convergence rates as the Legendre PS method numerical solutions throughout the displayed range of N values.

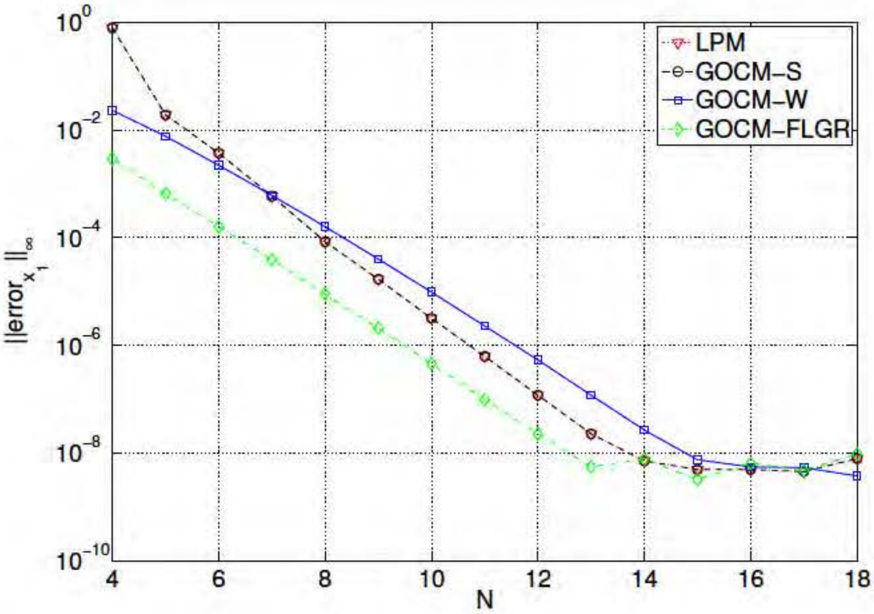


Figure 17: State x_1 approximation error vs. polynomial order, N, for Example 7.1.

However, the GOCM-W and GOCM-FLGR numerical solutions of x_2 and u show a marked improvement over that of the GOCM-S. Observations from this example show that the GOCM-W and GOCM-FLGR formulations have the potential to provide increased accuracies of the control solution with lower order approximations.

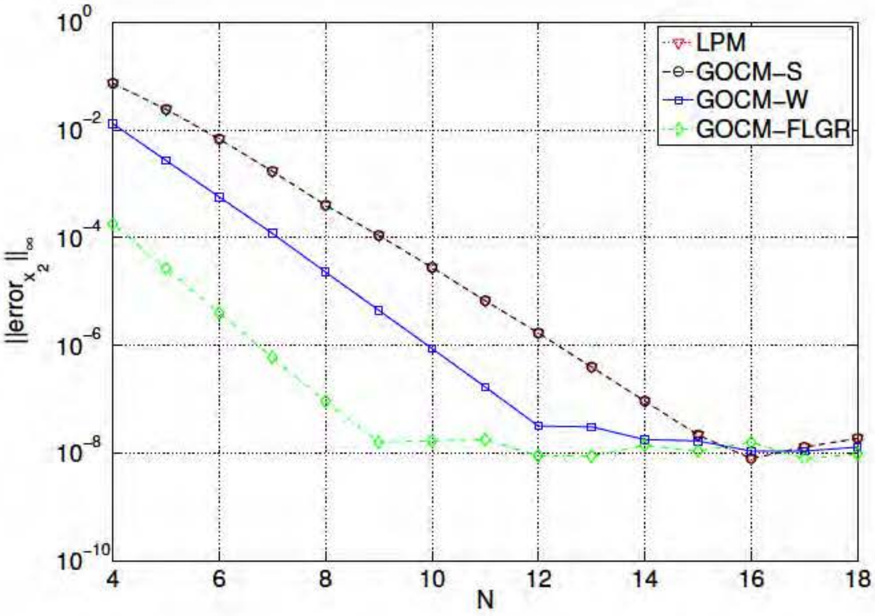


Figure 18: State x_2 approximation error vs. polynomial order, N, for Example 7.1.

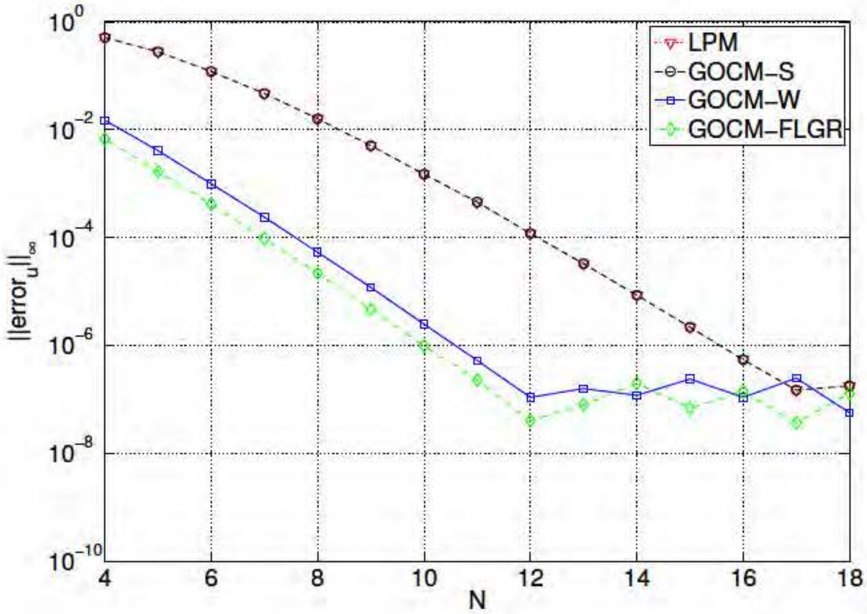


Figure 19: Control u approximation error vs. polynomial order, N, for Example 7.1.

7.2. Example 7.2: Nonlinear Two-Dimensional Problem with Fixed Boundary Conditions

Consider the *nonlinear* two-dimensional problem with *fixed boundary conditions* given by Kang [6] of minimizing the cost function

$$J = \int_0^{\pi} (1 - x_1 + x_1 x_2 + x_1 u)^2 dt, \tag{7.2}$$

subject to the dynamics

$$\dot{x}_1(t) = -x_1^2 x_2 \quad \text{and} \quad \dot{x}_2(t) = -1 + \frac{1}{x_1} + x_2 + \sin t + u,$$

and with boundary conditions

$$x_1(0) = 1, \quad x_2(0) = 0, \quad x_1(\pi) = \frac{1}{\pi + 1} \quad \text{and} \quad x_2(\pi) = 2.$$

The analytic solution to this problem is given by

$$\begin{aligned} x_1(t) &= \frac{1}{1 - \sin t + t}, \\ x_2(t) &= 1 - \cos t, \\ u(t) &= -(t + 1) + \sin t + \cos t, \end{aligned}$$

obtained via Pontryagin's maximum principle. This problem was solved using the GOCM-W (Section 5.1) with optimality and feasibility tolerances of 10^{-8} and 10^{-8}, respectively. A two-point initial guess was provided. Figure 20 shows a comparison of the exact solution with the GOCM-W approximation and $N = 20$.

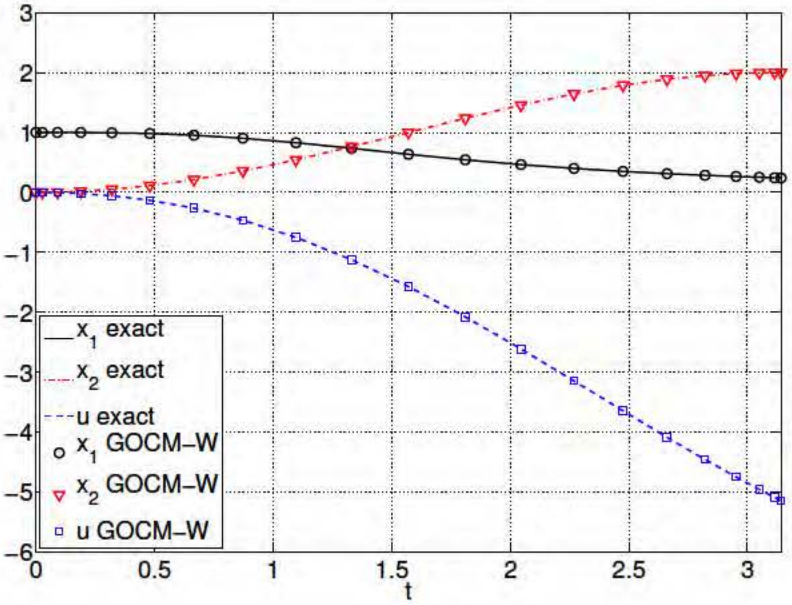


Figure 20: Exact solution and GOCM-W approximation with $N = 20$ for Example 7.2.

Figures 21, 22 and 23 show the maximum error of the states and control,

$$\begin{aligned}
\|\text{error}_{x_1}\|_\infty &= \left\|x_1(t_i) - x_1^N(t_i)\right\|_\infty, \quad i = 0, 1, \ldots, N,\\
\|\text{error}_{x_2}\|_\infty &= \left\|x_2(t_i) - x_1^N(t_i)\right\|_\infty, \quad i = 0, 1, \ldots, N,\\
\|\text{error}_{u}\|_\infty &= \left\|u(t_i) - u^N(t_i)\right\|_\infty, \quad i = 0, 1, \ldots, N,
\end{aligned}$$

respectively, for the LPM (Legendre PS method, Section 3.2), GOCM-S (Chapter 4), GOCM-W and GOCM-OI (Section 5.5.1) vs. polynomial order, N. Note that the GOCM-W approximation represents a formulation for which the initial conditions are enforced weakly and the final conditions are imposed in a strong sense through an endpoint constraint. Although enforcing all boundary conditions weakly provides accurate solutions, a *partial* weak enforcement of end conditions produces higher accuracies. Also note that the GOCM-OI approximations represent the over integration of this GOCM-W formulation. The GOCM-S approximations show nearly the exact same exponential convergence rates as the Legendre PS method numerical solutions throughout the displayed range of N values.

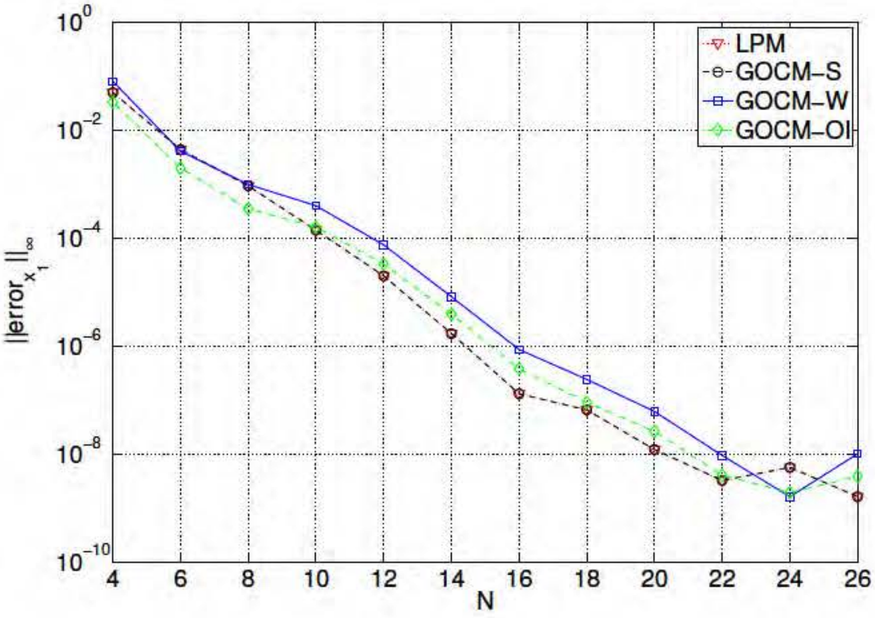


Figure 21: State x_1 approximation error vs. polynomial order, N, for Example 7.2.

The GOCM-W numerical solutions of x_2 and u show a marked improvement over that of the GOCM-S for $N \leq 22$; the control error for the GOCM-W is up to two orders of magnitude lower than that of the GOCM-S for the N values in this range. Finally, the GOCM-OI formulation has a smoothing effect on the accuracies of the GOCM-W approximations. While the GOCM-OI approximations of x_2 appear to be less superior to that of the GOCM-W, the GOCM-OI approximations of u gain slightly in accuracy for the lower order approximations. Observations from this example show that the GOCM-W and GOCM-OI formulations have the potential to provide increased accuracies of the control solution with lower order approximations.

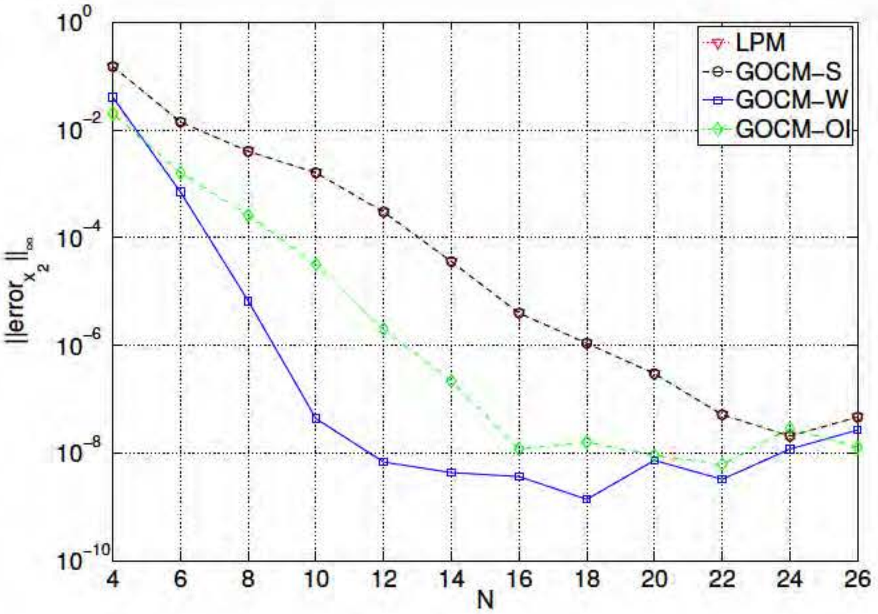


Figure 22: State x_2 approximation error vs. polynomial order, N, for Example 7.2.

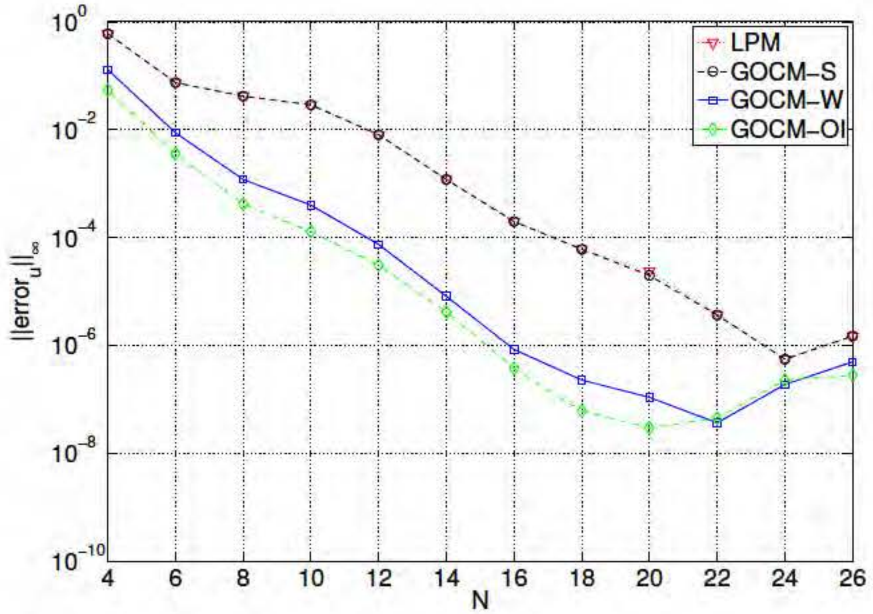


Figure 23: Control u approximation error vs. polynomial order, N, for Example 7.2.

7.3. Example 7.3: Two-Dimensional Bang-Bang Control Problem with Fixed Boundary Conditions

Consider the 2-dimensional *bang-bang control* problem with *fixed initial conditions* given by Pinch [21] of controlling the system from the initial point (a, b) at $t = 0$ to the origin $(0, 0)$ in the shortest time. Here, we minimizing the cost function

$$J = \int_0^{t_f} dt = t_f,$$

subject to the dynamics

$$\dot{x}(t) = y(t) \quad \text{and} \quad \dot{y}(t) = u(t),$$

and with boundary conditions

$$x(0) = a, \quad y(0) = b, \quad x(t_f) = 0, \quad y(t_f) = 0,$$

and control constraint $|u(t)| \leq k$. For the case where $a = 1$, $b = 3$ and $k = 1$, the analytic solution to this problem is given by

$$x(t) = \begin{cases} -\frac{t^2}{2} + 3t + 1, & t \leq t_\xi, \\ \frac{t^2}{2} - t_f t + \frac{t_f^2}{2}, & t > t_\xi, \end{cases}$$

$$y(t) = \begin{cases} -t + 3, & t \leq t_\xi, \\ t - t_f, & t > t_\xi, \end{cases},$$

$$u(t) = \begin{cases} -1, & t \leq t_\xi, \\ 1, & t > t_\xi, \end{cases}$$

obtained via Pontryagin's maximum principle, where $t_\xi = 3 + \sqrt{\frac{11}{2}}$ and $t_f = 3 + 2\sqrt{\frac{11}{2}}$. This problem was solved using the GOCM-S (Chapter 4), GOCM-W (Section 5.1), GOCM-

CG (Section 5.2) and GOCM-DG (Section 5.3). A two-point initial guess was provided for each approximation.

Figures 24 and 25 show comparisons of the exact solution with the GOCM-S and GOCM-W numerical solutions, respectively, with approximation order, $N = 20$.

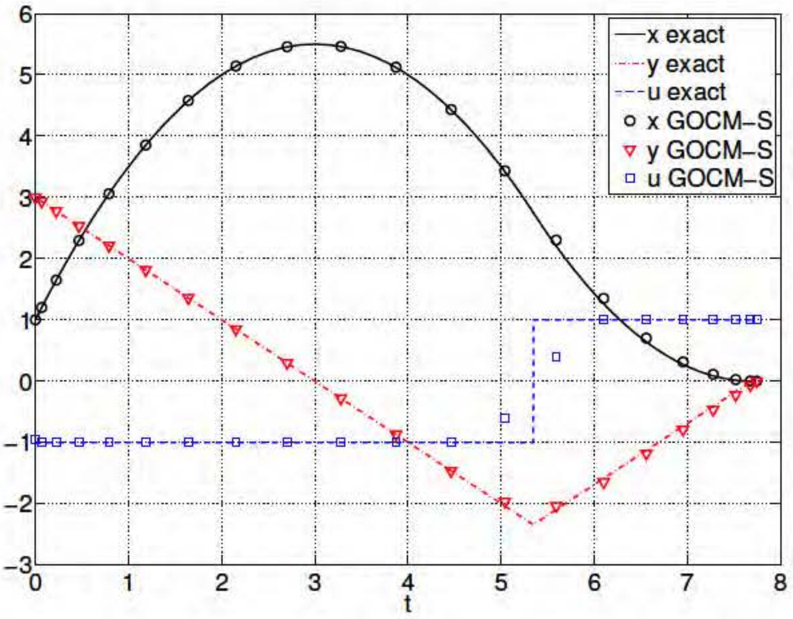


Figure 24: Exact solution and GOCM-S approximation with $N = 20$ for Example 7.3.

From observations, we can see a slight improvement from the GOCM-S to the GOCM-W numerical solutions, particularly in the approximation of the control, u, in the vicinity of the discontinuity location, t_{ξ}. However, the GOCM-S and GOCM-W have difficulty in approximating the discontinuous control solution. Even with an increased approximation order, there remains a maximum error of $O(10^{-1})$ for the GOCM-S and GOCM-W approximations of the control.

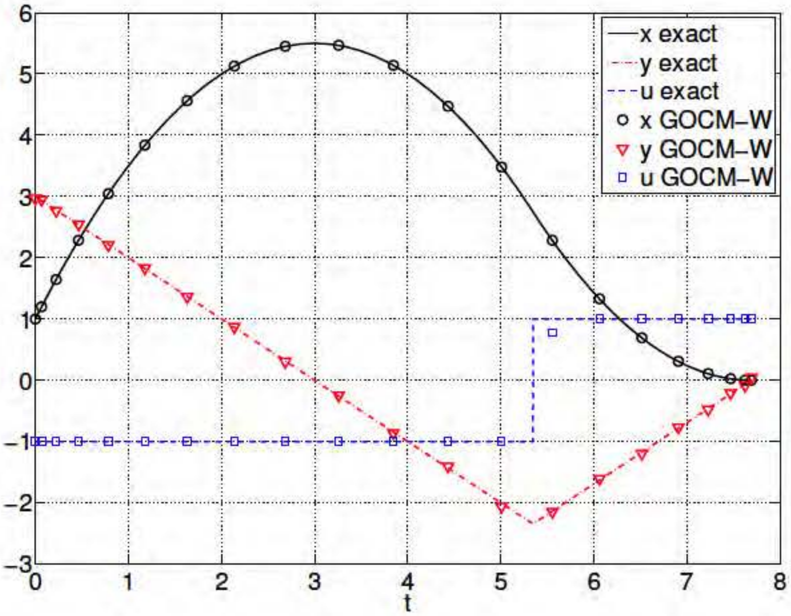


Figure 25: Exact solution and GOCM-W approximation with $N = 20$ for Example 7.3.

Figures 26 and 27 show comparisons of the exact solution with the GOCM-CG and GOCM-DG approximations, respectively, with number of *nonuniform* elements, $N_e = 2$, polynomial order inside each element, $N = 10$, and the boundary of the two elements located at t_ξ. It should be noted that in the element based formulations, t_ξ is defined as a decision variable in the NLP. An initial guess for t_ξ is then provided to the NLP with bounds prescribed. The total number of points for the GOCM-CG approximation is, $N_p = (N_e N + 1) = 21$ while for the GOCM-DG approximation, $N_p = (N + 1)N_e = 22$.

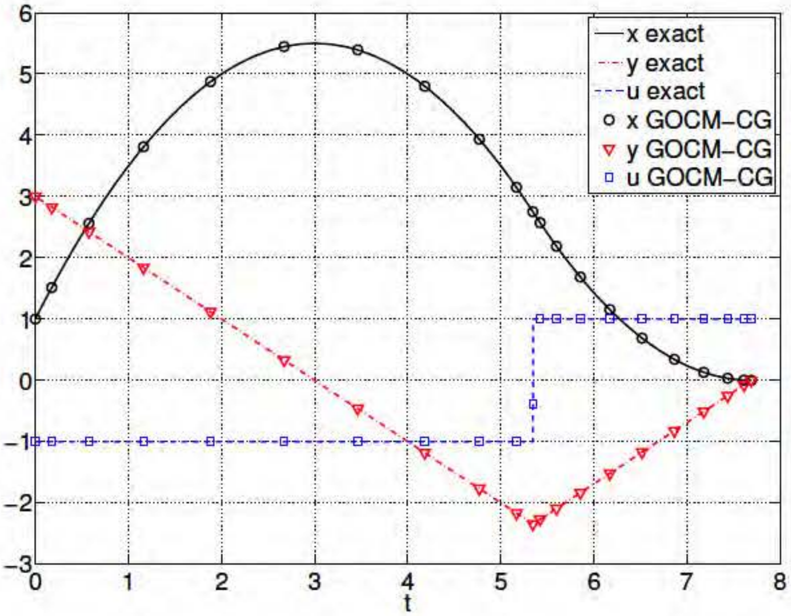


Figure 26: Exact solution and GOCM-CG approximation with $N_p = 21$, $N_e = 2$ and the boundary of the elements located at t_ξ for Example 7.3.

Note that both the GOCM-CG and GOCM-DG approximations of the states, x and y, achieve maximum errors of $O(10^{-9})$ from the exact states. Additionally, the GOCM-DG achieves an impressive $O(10^{-8})$ maximum error from the exact control, u, as compared with an accuracy of $O(10^{-1})$ for the GOCM-CG.

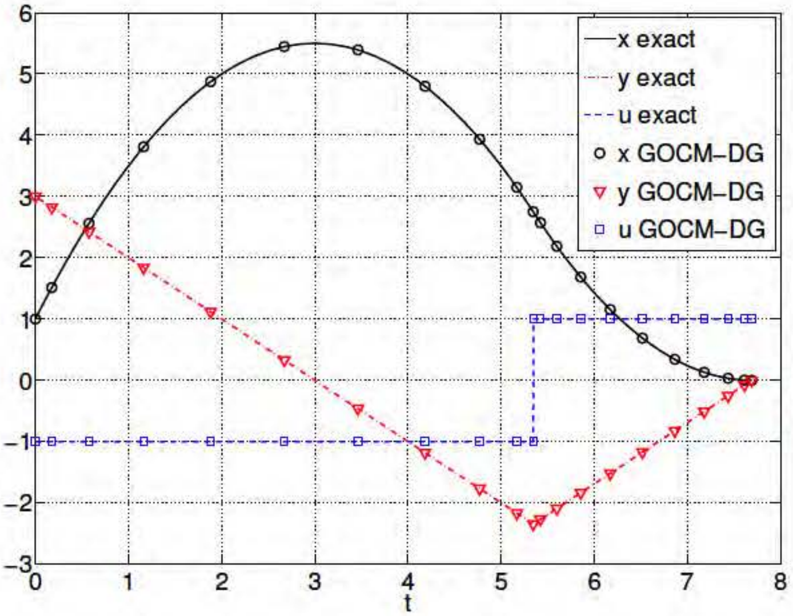


Figure 27: Exact solution and GOCM-DG approximation with $N_p = 22$, $N_e = 2$ and the boundary of the elements located at t_ξ for Example 7.3.

7.4. Example 7.4: Two-Dimensional Problem with Fixed Boundary Conditions

Consider again the *linear* two-dimensional problem with *fixed boundary conditions* given in Example 3.1 of minimizing the cost function

$$J = \frac{1}{2}\int_0^{t_f} u^2 dt, \tag{7.3}$$

subject to the dynamics

$$\dot{x}_1(t) = x_2 \quad \text{and} \quad \dot{x}_2(t) = C\sin(kt) + u,$$

and with boundary conditions

$$x_1(0) = 0, \quad x_2(0) = 0, \quad x_1(t_f) = 1 \quad \text{and} \quad x_2(t_f) = 0.$$

The analytic solution to this problem is given by Equations (3.9)–(3.13), where $t_f = 10$, $C = 0.1$ and $k = 8$. This problem was solved using the GOCM-W (Section 5.1) with optimality and feasibility tolerances of 10^{-8} and 10^{-8}, respectively. A two-point initial guess was provided for each approximation.

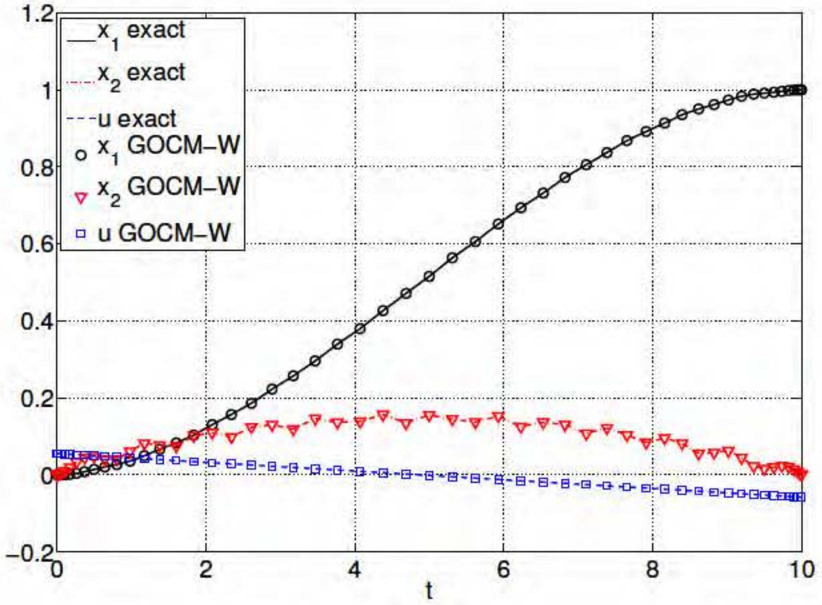


Figure 28: Exact solution and GOCM-W approximation with $N = 50$ for Example 7.4.

$$\|\text{error}_{x_1}\|_\infty = \left\|x_1(t_i) - x_1^N(t_i)\right\|_\infty, \quad i = 0, 1, \ldots, N,$$
$$\|\text{error}_{x_2}\|_\infty = \left\|x_2(t_i) - x_1^N(t_i)\right\|_\infty, \quad i = 0, 1, \ldots, N,$$
$$\|\text{error}_{u}\|_\infty = \left\|u(t_i) - u^N(t_i)\right\|_\infty, \quad i = 0, 1, \ldots, N,$$

respectively, for the LPM (Legendre PS method, Section 3.2), GOCM-S (Chapter 4), GOCM-W and GOCM-LG (Section 5.5.2.2) vs. polynomial order, N. Note that, as with Example 7.2, the GOCM-W approximation represents a formulation for which the initial conditions are enforced weakly and the final conditions are imposed in a strong sense through an endpoint constraint. Although enforcing all boundary conditions weakly provides accurate solutions, a *partial* weak enforcement of end conditions produces higher accuracies. Note that the GOCM-S approximations show nearly the exact same exponential convergence rates as the LPM numerical solutions throughout the displayed range of N values.

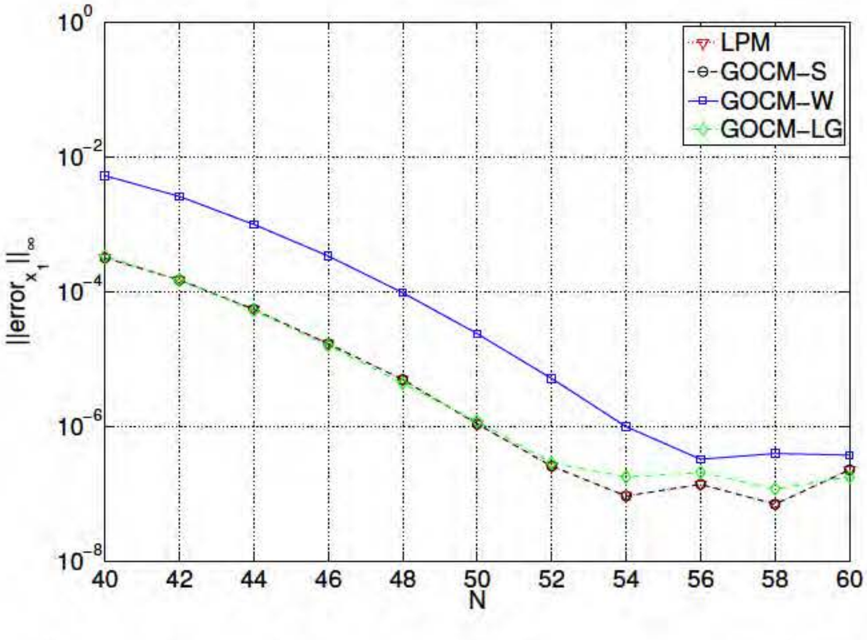


Figure 29: State x_1 approximation error vs. polynomial order, N, for Example 7.4.

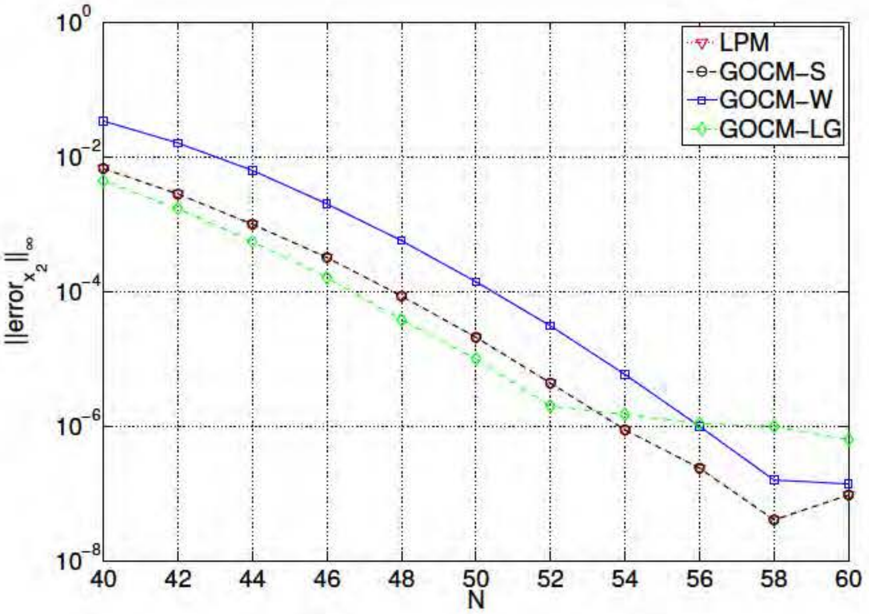


Figure 30: State x_2 approximation error vs. polynomial order, N, for Example 7.4.

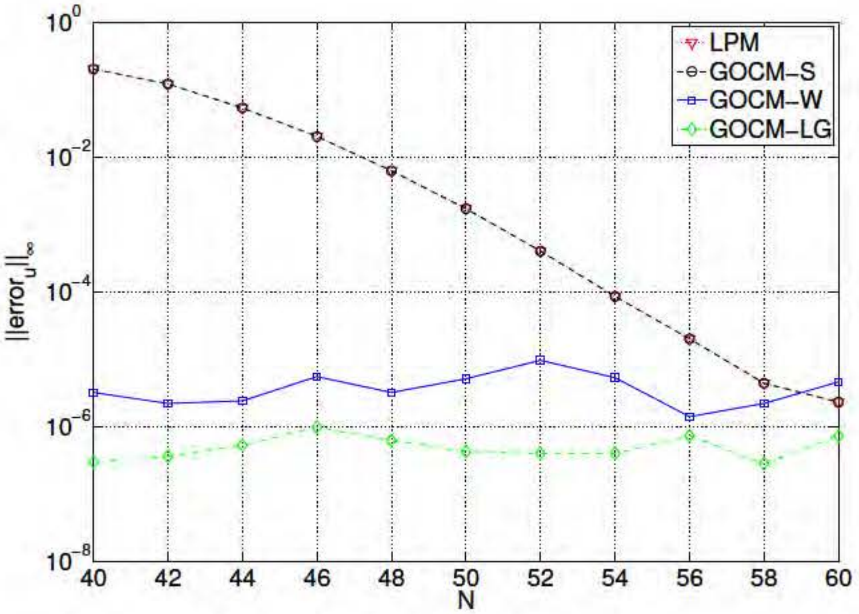


Figure 31: Control u approximation error vs. polynomial order, N, for Example 7.4.

However, the GOCM-W and GOCM-LG numerical solutions for u show a marked improvement over that of the GOCM-S, particularly in the lower order approximations; the control error for the GOCM-LG is one to five orders of magnitude lower than that of the GOCM-S in this range. Observations from this example show that the GOCM-W and GOCM-LG formulations have the potential to provide increased accuracies of the control solution with lower order approximations.

7.5. Example 7.5: Two-Dimensional Multi-scale Problem with Fixed Boundary Conditions

Consider again the *linear* two-dimensional problem with *fixed boundary conditions* given in Example 3.1 and 7.4. Here we take advantage of the multi-scale nature of the problem. The slow state, x_1, fast state, x_2 and control u are discretized on LGL grids, $\{\tau_j\}_{j=0}^{N_{x_1}}$, $\{t_j\}_{j=0}^{N_{x_1}}$ and $\{\tilde{\tau}_j\}_{j=0}^{N_u}$, respectively, where $N_{x_1}, N_u < N_{x_2}$. This problem was solved with optimality and feasibility tolerances of 5×10^{-4} and 5×10^{-3}, respectively. The exact solution was used as an initial guess. Figure 32 shows a comparison of the exact solution with the GOCM-MS (Section 5.4) numerical solutions with $N_{x_1} = 10$, $N_{x_2} = 50$ and $N_u = 5$.

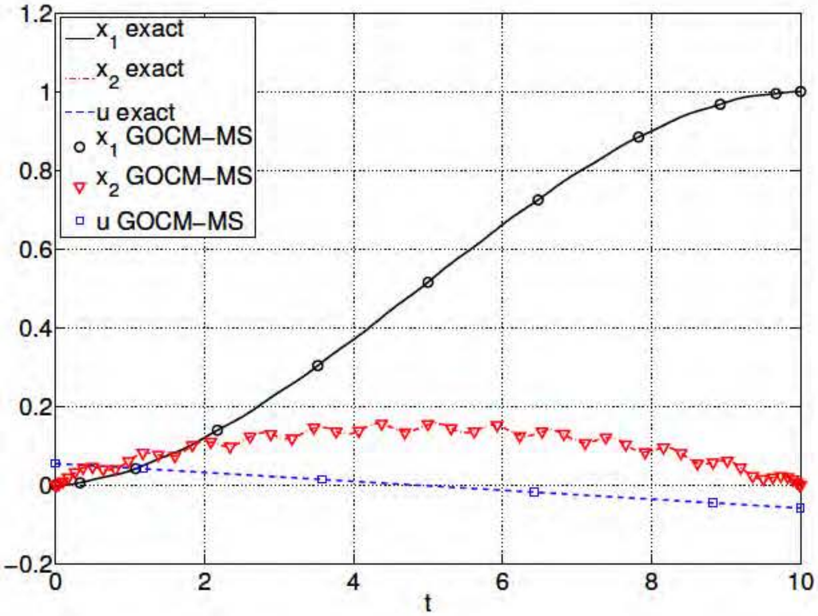


Figure 32: Exact solution and GOCM-MS approximation with $N_{x_1} = 10$, $N_{x_2} = 50$ and $N_u = 5$ for Example 7.5.

Unlike the multi-scale Legendre PS method outlined in Section 3.4, the GOCM-MS provides a feasible solution. In fact, the maximum error of the GOCM-MS numerical solutions all have magnitude $O(10^{-4})$. It is also important to point out that the required feasibility tolerance of 5×10^{-3} is the same order of magnitude as the largest x_1 spectral coefficient dropped from the approximation (see Figure 13). This observation is consistent with Remark 3.6. Additionally, Remark 3.6 highlights the potential for a feasible solution with the lower bounds: $3 \leq N_{x_1}$, $43 \leq N_{x_2}$ and $1 \leq N_u$ (note the associated magnitudes of the Legendre spectral coefficients in Figure 13 for x_1, x_2 and u). This is in fact the case: Figure 33 shows a comparison of the exact solution with the GOCM-MS numerical solutions with $N_{x_1} = 3$, $N_{x_2} = 43$ and $N_u = 1$, solved with optimality and feasibility tolerances of 5×10^{-4} and 5×10^{-3}, respectively.

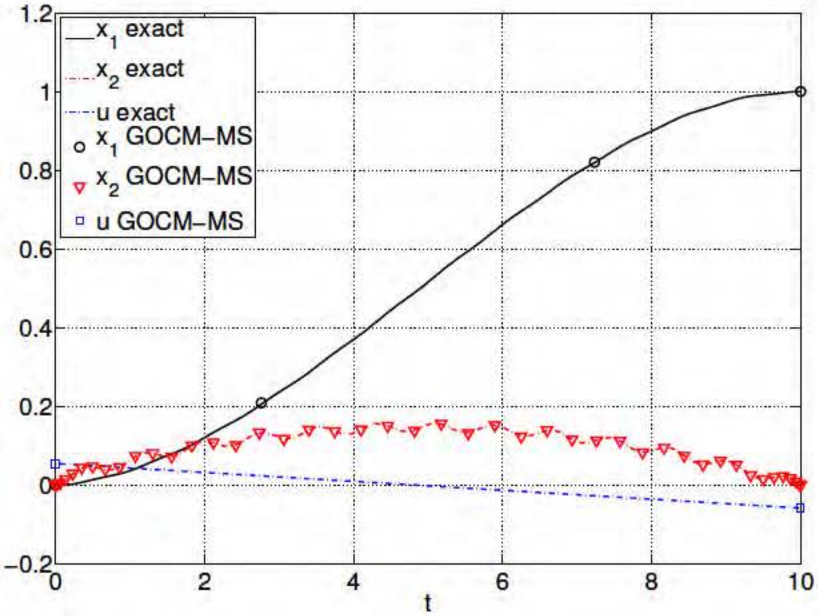


Figure 33: Exact solution and GOCM-MS approximation with $N_{x_1} = 3$, $N_{x_2} = 43$ and $N_u = 1$ for Example 7.5.

The maximum errors of the new GOCM-MS approximation (with $N_{x_1} = 3$, $N_{x_2} = 43$ and $N_u = 1$) remain at $O(10^{-4})$ for the control, u, and $O(10^{-3})$ for the states, x_1 and x_2. Note that when the full-scale approach was used in Example 7.4, a GOCM-W approximation order of $N > 40$ was required for the same order of accuracies. It is clear that the size of the NLP for this problem has been reduced significantly. The number of decision variables needed for the multi-scale Galerkin optimal control formulation is nearly 33% of that required of the full-scale problem solved with the GOCM-W. Additionally, this multi-scale approach has the potential to more efficiently solve a great number of optimal control problems. The larger the dimension of the multi-scale problem, the greater the potential savings in computational efficiency!

7.6. Example 7.6: Nonlinear Two-Dimensional Multi-scale Problem with Fixed Initial Conditions

Consider the following *nonlinear* two-dimensional problem with *fixed initial conditions* given by Gong et al. [71] of minimizing the cost function

$$J = \int_0^1 (x_1 - t)^2 dt, \tag{7.4}$$

subject to the dynamics

$$\dot{x}_1(t) = \sin(50x_1) + x_2 \quad \text{and} \quad \dot{x}_2(t) = u, \tag{7.5}$$

and with initial conditions

$$x_1(0) = 0 \quad \text{and} \quad x_2(0) = 1. \tag{7.6}$$

The analytic solution to this problem is given by

$$\begin{aligned} x_1(t) &= t, \\ x_2(t) &= 1 - \sin(50t), \\ u(t) &= -50\cos(50t), \end{aligned}$$

obtained via Pontryagin's maximum principle. Figures 34 and 35 show a comparison of the exact state and control solutions with the GOCM-W (Section 5.1) numerical solutions, respectively, with $N = 45$.

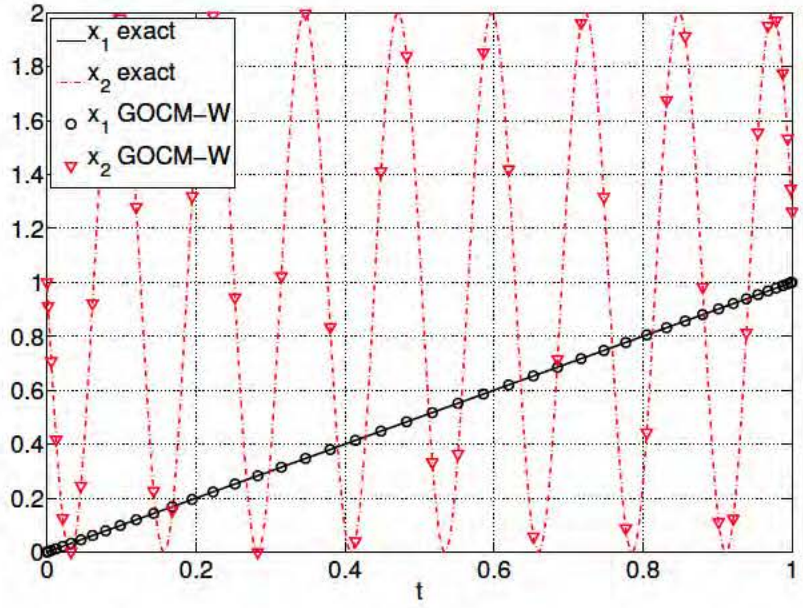


Figure 34: Exact state solutions and GOCM-W approximation with $N = 45$ for Example 7.6.

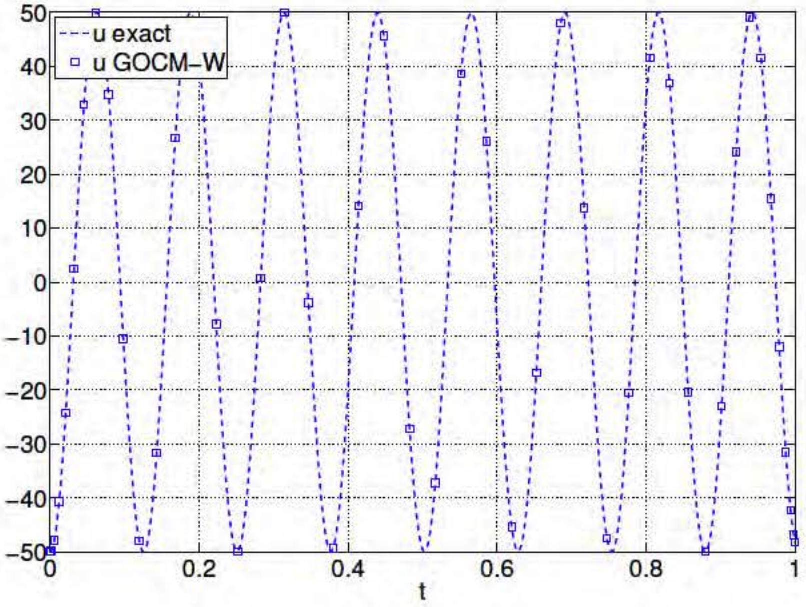
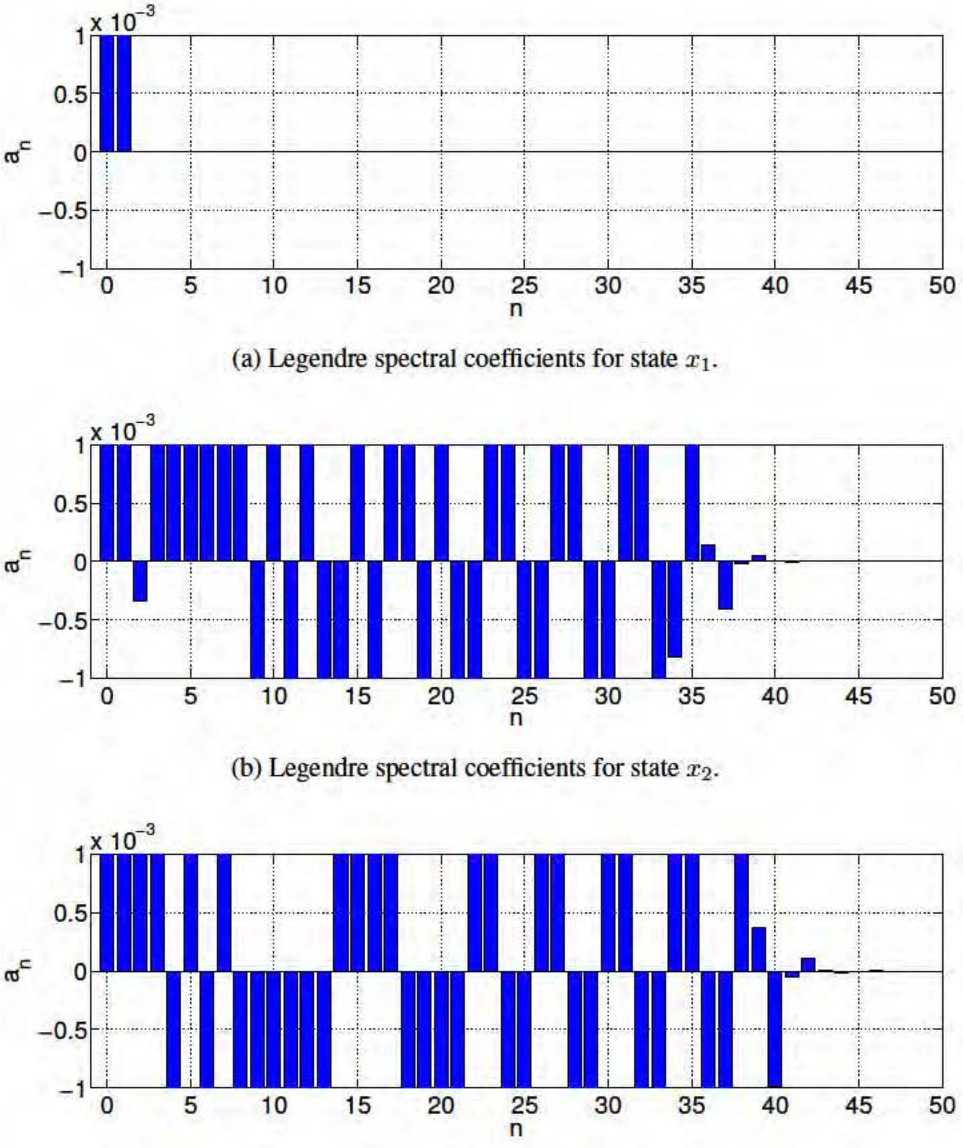


Figure 35: Exact control solution and GOCM-W approximation with $N = 45$ for Example 7.6.

It is clear from Figures 34 and 35 that problem (7.4)–(7.6) may be considered multiscale, where x_1 is the slow state and x_2 the fast state. Consider this problem now recast in the form of Problem $\tilde{\text{B}}$. Here, we discretize the slow state, x_1, fast state, x_2, and control, u, on LGL grids, $\{\tau_j\}_{j=0}^{N_{x_1}}$, $\{t_j\}_{j=0}^{N_{x_2}}$ and $\{\tilde{\tau}_j\}_{j=0}^{N_u}$, respectively. In order to determine N_{x_1}, N_{x_2} and N_u we consider the spectral coefficients of x_1, x_2 and u given in Figure 36.



(a) Legendre spectral coefficients for state x_1.

(b) Legendre spectral coefficients for state x_2.

(c) Legendre spectral coefficients for control u.

Figure 36: Legendre spectral coefficients for x_1, x_2 and u for Example 7.6.

If the feasibility tolerance is set to 10^{-3}, Remark 3.6 highlights the potential for a feasible solution with the bounds: $1 \leq N_{x_1}$, $39 \leq N_{x_2}$ and $42 \leq N_u$. This is in fact the case: Figures 37 and 38 show comparisons of the exact states and control with the GOCM-MS (Section 5.4) numerical solutions with $N_{x_1} = 1$, $N_{x_2} = 39$ and $N_u = 42$, solved with optimality and feasibility tolerances of 10^{-4} and 10^{-3}, respectively.

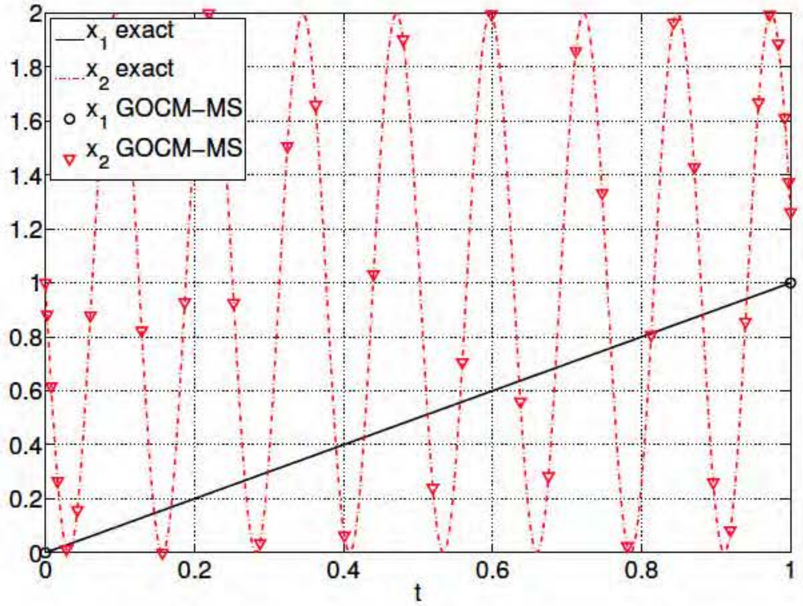


Figure 37: Exact state solutions and GOCM-MS approximation with $N_{x_1} = 1$ and $N_{x_2} = 39$ for Example 7.6.

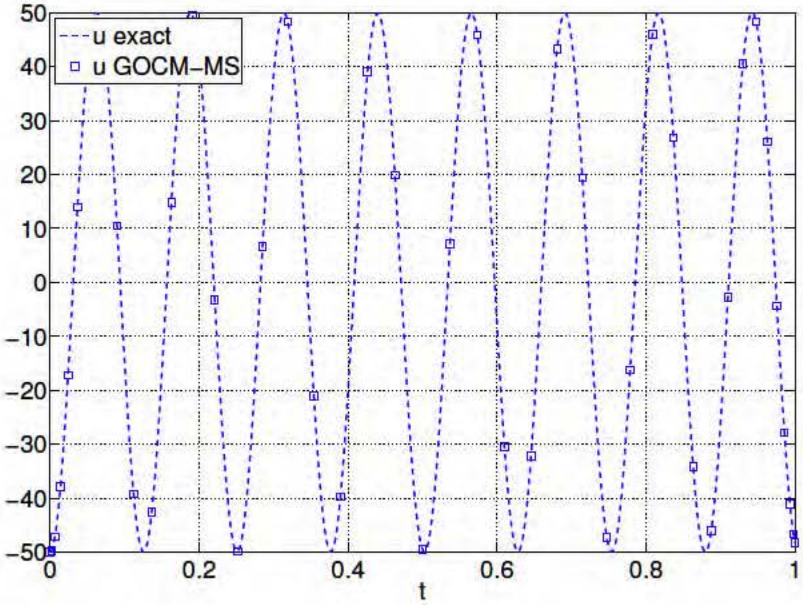


Figure 38: Exact control solution and GOCM-MS approximation with $N_u = 42$ for Example 7.6.

The maximum errors of the new GOCM-MS approximation (with $N_{x_1} = 1$, $N_{x_2} = 39$ and $N_u = 42$) are $O(10^{-3})$ for the control, u, and $O(10^{-6})$ and $O(10^{-4})$ for states, x_1 and x_2, respectively.

7.7. Summary

Galerkin optimal control is a versatile family of numerical formulations for solving optimal control problems. The examples shown in this chapter demonstrate the potential of this Galerkin-based family of formulations outlined in Chapters 4–6. Three particular highlights of Galerkin optimal control is its ability to weakly enforce problem end conditions, handle problems with discontinuous solutions and its potential to reduce the size of multi-scale problems.

Examples 7.1, 7.2 and 7.4 demonstrate the improved accuracy provided by the Galerkin optimal control formulation with weak enforcement of boundary conditions over strong enforcement for both boundary value problems as well as problems with incomplete sets of end conditions. In particular, the GOCM-W shows an advantage over the GOCM-S for low order approximations of control solutions. Examples 7.1 and 7.4 also show the potential advantages of the Galerkin formulations with F-LGR and LG points, respectively. Example 7.3 demonstrates the effectiveness of the element-based Galerkin optimal control formulations (such as the GOCM-DG) when employed to approximated optimal control problems with discontinuous controls. Lastly, Examples 7.5 and 7.6 demonstrate the computational efficiency in which the multi-scale Galerkin optimal control formulation may solve multi-scale problems, those with states and controls that evolve on different timescales. In contrast to the difficulties with the multi-scale Legendre PS method (see Section 3.4) highlighted in Chapter 3, the GOCM-MS is shown to successfully reduce the size of multi-scale problems.

CHAPTER 8: CONCLUSIONS AND AREAS FOR FUTURE RESEARCH

8.1. Dissertation Summary

This dissertation introduced and developed the theory for a Galerkin-based family of numerical formulations that calculate optimal trajectories by discretizing the system dynamics using Galerkin numerical techniques and approximate the cost function with Gaussian quadrature. An important result of the Galerkin formulations are that they can be used to prove feasibility and consistency theorems that apply to optimal control problems with continuous and/or piecewise continuous controls. It was shown that Galerkin optimal control may be formulated in a variety of ways to allow for efficiency and/or improved accuracy while solving a wide range of optimal control problems.

A highlight of Galerkin optimal control is its ability to be formulated to enforce boundary conditions in a weak sense, imposing end conditions only up to the accuracy of the numerical approximation itself. The increased approximation accuracy of the weak boundary formulation (particularly in the approximation of control solutions) was shown on several linear and nonlinear problems. It was also demonstrated that the Galerkin optimal control formulation with weak imposition of end conditions allows for problem discretizations with other than LGL points. Galerkin optimal control with Legendre-Gauss-Radau and Legendre-Gauss points were shown to be advantageous due to the increased accuracy of solutions. Galerkin optimal control may also be formulated with other than Lagrangian test functions, such as the Legendre polynomials.

In addition, Galerkin optimal control has proven to be effective in reducing the dimension of multi-scale problems, those in which states and controls evolve on different timescales. In one example presented the number of decision variables required by the multi-scale Galerkin optimal control formulation was nearly 33% of that required of the full-scale problem. The multi-scale formulation has the potential to more efficiently solve a great number of optimal control problems, certainly those that include fast and slow dynamics.

For optimal control problems with discontinuities (such as bang-bang control problems), an element-based approach has shown to be beneficial. In general, using the element-based Galerkin optimal control formations discussed (both continuous and discontinuous Galerkin) may lead to higher computational efficiencies and the formulation may be retooled to incorporate hp-adaptive techniques, such as the spectral algorithm discussed in [86]. Additionally, the discontinuous Galerkin formulation may be advantageous from a parallel computing standpoint.

Galerkin optimal control has demonstrated exponential convergence for a large class of problems. It is clear that Galerkin optimal control is a versatile and accurate family of formulations that has the potential to provide real time optimal control solutions for a number of applications.

8.2. Future Work

Galerkin optimal control shows the potential for solving a wide range of optimal control problems with a variety of state and control constraints. However, application of the Galerkin optimal control formulations to real-world problems have been somewhat limited due to research time limitations. Future application of Galerkin optimal control to problems with real-world conditions is necessary to demonstrate its true versitiliy. Testing Galerkin optimal control on different types of control problems will inevitably highlight the strengths (and weaknesses) of each formulation. Lastly, this dissertation includes a number of important theorems that serve as the theoretical foundations for Galerkin optimal control, however, the list is not complete. Future work to increase the theoretical underpinnings of Galerkin optimal control, to include a rate of convergence analysis, is forthcoming.

APPENDIX A:
NORMS AND FUNCTIONAL SPACES

Throughout this dissertation, the L^p spaces, for $1 \leq p < \infty$, are used quite often. They consist of all measurable functions $v : [-1, 1] \to \mathbb{R}$, such that [87]

$$\int_{-1}^{1} |v(t)|^p dt < \infty,$$

with finite L^p-norm,

$$\|v\|_{L^p} = \left(\int_{-1}^{1} |v(t)|^p dt \right)^{\frac{1}{p}} < \infty.$$

In particular, the L^2 space is referenced quite readily, defined by

$$\|v\|_{L^2} = \left(\int_{-1}^{1} |v(t)|^2 dt \right)^{\frac{1}{2}},$$

which is induced by the inner product

$$(u, v) = \int_{-1}^{1} u(t)v(t)dt.$$

Additionally, the L^∞ space consist of all measurable functions $v : [-1, 1] \to \mathbb{R}$ such that $|v(t)| \leq M$ for almost all $t \in [-1, 1]$. The L^∞-norm can be expressed as

$$\|v\|_{L^\infty} = \inf\{M||v(t)| \leq M \text{ almost everywhere on } t \in [-1, 1]\}.$$

The idea of a function of bounded variation is also used within this dissertation. To frame this idea first we must define the total variation, $V(u)$. For a function $u : [-1, 1] \to \mathbb{R}$

the total variation of u on $[-1,1]$ is defined as

$$V(u) = \sup\{\sum_{j=1}^{N} |u(t_j) - u(t_{j-1})| \ \{t_k\}_{k=0}^{N} \in P\},$$

where P is the set of all finite partitions of [-1,1]. If the total variation is bounded, $V(u) < \infty$, then u is said to be of bounded variation in $[-1,1]$.

APPENDIX B: DISCRETE NORMS

Let grid $\{t_j\}_{j=0}^N$ be associated with quadrature weights $\{w_j\}_{j=0}^N$ (see Section 2.2.2). Then the discrete norm, $\|v\|_N$, from [46] is defined as the quantity

$$\|v\|_N = (v, v)_N^{\frac{1}{2}},$$

where the discrete inner product is accomplished by numerical integration and given by

$$(v, v)_N = \sum_{j=0}^{N} v(t_j)^2 w_j.$$

In the case that $v^2 \in P_{2N+\delta}$ and $\{t_j\}_{j=0}^N$ are LG, LGR or LGL points, the numerical integration is exact and

$$(v, v)_N = \int_{-1}^{1} u^2 dt,$$

where $\delta = 1, 0, -1$ for LG, LGR or LGL points, respectively, and $\{w_j\}_{j=0}^N$ are the associated weights.

Additionally, there exist constants $\alpha, \beta > 0$ such that

$$\alpha\|v\|_{L^2} \leq \|v\|_N \leq \beta\|v\|_{L^2}$$

for all $v \in P_N$.

Lastly, $\|v\|_\infty$ denotes the maximum element of vector, $v \in \mathbb{R}^n$.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX C: SOBOLEV SPACES

Throughout this dissertation, the Sobolev spaces, $W^{m,p}$, are referenced quite often. They consist of all functions, $v : [-1, 1] \to \mathbb{R}^n$ having weak derivative, $v^{(i)} \in L^p$, where $0 \leq i \leq m$, with the norm [87]

$$\|v\|_{W^{m,p}} = \left(\sum_{i=0}^{m} \left\| v^{(i)} \right\|_{L^p}^{p} \right)^{\frac{1}{p}},$$

where $\|v\|_{L^p}$ denotes,

$$\|v\|_{L^p} = \left(\int_{-1}^{1} |v(t)|^p dt \right)^{\frac{1}{p}}.$$

The seminorm may be expressed as

$$|v|_{W^{m,p;N}} = \left(\sum_{i=\min(m,N+1)}^{m} \left\| v^{(i)} \right\|_{L^p}^{p} \right)^{\frac{1}{p}}.$$

Additionally, Sobolev spaces may be defined with a fractional order. They consist of all measurable functions $v : [-1, 1] \to \mathbb{R}$, such that [88]

$$W^{\sigma,p} = \{ v \in L^p : \int_{-1}^{1} \int_{-1}^{1} \frac{|v(x) - v(y)|^p}{|x - y|^{1+\sigma p}} dx dy < \infty \},$$

for $0 < \sigma < 1$ and $1 \leq p < \infty$, with the norm,

$$\|v\|_{W^{\sigma,p}} = \left(\int_{-1}^{1} |v(t)|^p dt + \int_{-1}^{1} \int_{-1}^{1} \frac{|v(x) - v(y)|^p}{|x - y|^{1+\sigma p}} dx dy \right)^{\frac{1}{p}}.$$

Lastly, due to the extensive use of the space $W^{m,2}$, notation is simplified by letting $W^{m,2} = H^m$.

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

[1] G. N. Elnagar, M. A. Kazemi, and M. Razzaghi, “The pseudospectral legendre method for discretizing optimal control problems,” *IEEE Transactions on Automatic Control*, vol. 40, pp. 1793–1796, Oct. 1995.

[2] F. Fahroo and I. M. Ross, “Costate estimation by a legendre pseudospectral method,” in *Proceedings of the 1998 AIAA Guidance, Navigation and Control Conference*, no. AIAA 1998-4222, (Boston, MA), AIAA, Aug. 1998.

[3] Q. Gong, W. Kang, and I. M. Ross, “A pseudospectral method for the optimal control of constrained feedback linearizable systems,” *IEEE Transactions on Automatic Control*, vol. 51, pp. 1115–1129, Jul. 2006.

[4] W. Kang and N. Bedrossian, “Pseudospectral optimal control theory makes debut flight, saves nasa $1m in under three hours,” *SIAM News*, vol. 40, pp. 1–3, Sep. 2007.

[5] Q. Gong, I. M. Ross, W. Kang, and F. Fahroo, “Connections between the covector mapping theorem and the convergence of pseudospectral methods for optimal control,” *Computational Optimization and Applications*, vol. 41, pp. 307–335, Dec. 2008.

[6] W. Kang, “Rate of convergence for a legendre pseudospectral optimal control of feedback linearizable systems,” *Journal of Control Theory and Applications*, vol. 8, pp. 391–405, Nov. 2010.

[7] J. Ruths and J.-S. Li, “A multidimensional pseudospectral method for optimal control of quantum ensembles,” *Journal of Chemical Physics*, vol. 134, p. 044128, Jan. 2011.

[8] N. Bedrossian, S. Karpenko, and S. Bhatt, “Overclock my satellite,” *IEEE Spectrum*, vol. 49, pp. 54–62, Nov. 2012.

[9] G. N. Elnagar and M. Razzaghi, “A collocation-type method for linear quadratic optimal control problems,” *Optimal Control Applications and Methods*, vol. 18, pp. 227–235, May/Jun. 1997.

[10] G. N. Elnagar and M. A. Kazemi, “Pseudospectral legendre-based optimal computation of nonlinear constrained variational problems,” *Journal of Computational and Applied Mathematics*, vol. 88, pp. 363–375, Mar. 1998.

[11] Q. Gong, I. M. Ross, and F. Fahroo, “A chebyshev pseudospectral method for nonlinear constrained optimal control problems,” in *Proceedings of the Joint 48th IEEE Conference on Decision and Control and 28th Chinese Control Conference*, no. ThBIn3.11, (Shanghai, P.R. China), IEEE, Dec. 2009.

[12] I. M. Ross and F. Fahroo, "Pseudospectral knotting methods for solving optimal control problems," *Journal of Guidance, Control, and Dynamics*, vol. 27, pp. 397–405, May-Jun. 2004.

[13] I. M. Ross, Q. Gong, and P. Sekhavat, "The bellman pseudospectral method," in *Preceedings of the 2008 AIAA/AAS Astrodynamics Specialist Conference and Exhibit*, no. AIAA 2008-6448, (Honolulu, HI), AIAA, Aug. 2008.

[14] B. Hulme, "One-step piecewise polynomial galerkin methods for initial value problems," *Mathematics of Computation*, vol. 26, pp. 415–426, Apr. 1972.

[15] B. Hulme, "Discrete galerkin and related one-step methods for ordinary differential equations," *Mathematics of Computation*, vol. 26, pp. 881–891, Oct. 1972.

[16] B. Cockburn, G. E. Karniadakis, and C.-W. Shu, *Discontinuous Galerkin Methods: Theory, Computation and Applications*, vol. 11 of *Lecture Notes in Computational Science and Engineering*, ch. The developement of discontinuous Galerkin methods, pp. 3–50. Berlin Heidelberg: Springer, 2000.

[17] J. H. Jellett, *An Elementary Treatise on the Calculus of Variations*. London: The University Press, William S. Orr and Co, 1850.

[18] J. J. O'Connor and E. F. Robertson, "The brachistochrone problem [online]," Feb. 2002. Available: http://www-history.mcs.st-andrews.ac.uk/HistTopics/Brachistochrone.html.

[19] H. J. Sussmann and J. C. Willems, "300 years of optimal control: from the brachystochrone to the maximum principle," *IEEE Control Systems Magazine*, vol. 17, pp. 32–44, Jun. 1997.

[20] E. T. Bell, *Men of Mathematics*. New York, NY: Simon and Schuster, Inc., 1937.

[21] E. R. Pinch, *Optimal Control and the Calculus of Variations*. New York: Oxford University Press Inc., 1993.

[22] L. Pontryagin, V. Boltyanskii, R. Gamkrelidze, and E. F. Mishchenko, *The Mathematical Theory of Optimal Processes*. New York, London: John Wiley and Sons, Inc., 1962.

[23] H. J. Pesch and M. Plail, "The maximum principle of optimal control: A history of ingenious ideas and missed opportunities," *Control and Cybernetics*, vol. 38, pp. 973–995, Dec. 2009.

[24] M. Helmut and H. J. Oberle, "Second order sufficient conditions for optimal control problems with free final time: the riccati approach," *SIAM Journal on Control and Optimization*, vol. 41, pp. 380–403, Jun. 2002.

[25] L. W. Neustadt, *Optimization: A Theory of Necessary Conditions*. Princeton, NJ: Princeton University Press, 1976.

[26] D. E. Kirk, *Optimal Control Theory: An Introduction*. Mineola, NY: Dover Publications, 2004.

[27] I. M. Ross, *A Primer on Pontryagin's Principle in Optimal Control*. Collegiate Publishing, 2009.

[28] R. F. Hartl, S. P. Sethi, and R. G. Vickson, "A survey of the maximum principles for optimal control problems with state constraints," *SIAM Review*, vol. 37, pp. 181–218, Jun. 1995.

[29] J. T. Betts, *Practical Methods for Optimal Control and Estimation Using Nonlinear Programming*, vol. 19 of *Advances in Design and Control*. 2nd ed. Philadelphia: SIAM, 2010.

[30] J. T. Betts, "Survey of numerical methods for trajectory optimization," *Journal of Guidance, Control, and Dynamics*, vol. 21, p. 193, Mar.-Apr. 1998.

[31] E. Trelat, "Optimal control and applications to aerospace: Some results and challenges," *Journal of Optimization Theory and Applications*, vol. 154, pp. 713–758, Sep. 2012.

[32] I. M. Ross and S. Karpenko, "A review of pseudospectral optimal control: From theory to flight," *Annual Reviews in Control*, vol. 36, pp. 182–197, Dec. 2012.

[33] O. von Stryk and R. Bulirsch, "Direct and indirect methods for trajectory optimization," *Annals of Operation Research*, vol. 37, pp. 357–373, Dec. 1992.

[34] P. Lu, H. Sun, and B. Tsai, "Closed-loop endoatmospheric ascent guidance," *Journal of Guidance, Control, and Dynamics*, vol. 26, pp. 283–294, Mar.-Apr. 2003.

[35] P. Williams, "Application of pseudospectral methods for receding horizon control," *Journal of Guidance, Control, and Dynamics*, vol. 27, pp. 310–314, Mar. 2004.

[36] F. Fahroo and I. M. Ross, "Advances in pseudospectral methods," in *Preceedings of the 2008 AIAA Guidance, Navigation, and Control Conference*, no. AIAA 2008-7309, (Honolulu, HI), AIAA, Aug. 2008.

[37] F. Fahroo and I. M. Ross, "Pseudospectral methods for infinite-horizon nonlinear optimal control problems," *Journal of Guidance, Control, and Dynamics*, vol. 31, pp. 927–936, Jul.-Aug. 2008.

[38] I. M. Ross, *A Beginner's Guide to DIDO: A MATLAB Application Package for solving Optimal Control Problems*. 7.3 ed. Monterey, CA, 2007.

[39] C. R. Hargraves and S. W. Paris, “Direct trajectory optimization using nonlinear programming and collocation,” *Journal of Guidance, Control, and Dynamics*, vol. 10, pp. 338–342, Jul. 1987.

[40] P. E. Gill, W. Murray, M. A. Saunders, and M. H. Wright, *User’s Guide for NPSOL 5.0: A FORTRAN Package for Nonlinear Programming*. 5th ed. Systems Optimization Laboratory, Department of Management Science and Engineering, Stanford University, Stanford, CA, June 2001.

[41] P. E. Gill, W. Murray, and M. A. Saunders, “Snopt: An sqp algorithm for large-scale constrained optimization,” *SIAM Review*, vol. 47, pp. 99–131, Jan. 2002.

[42] P. E. Gill, W. Murray, and M. A. Saunders, *User’s Guide for SNOPT Version 7: Software for Large-Scale Nonlinear Programming*. 7th ed. Systems Optimization Laboratory, Department of Management Science and Engineering, Stanford University, Stanford, CA, 2008.

[43] J. T. Betts and W. P. Huffman, “A sparse nonlinear optimization algorithm,” *Journal of Optimization Theory and ApplicationsTheory and Applications*, vol. 82, pp. 519–541, Sep. 1994.

[44] F. X. Giraldo, “Element-based galerkin methods.” class notes for MA4245: Mathematical Principles of Galerkin Methods, Department of Applied Mathematics, Naval Postgraduate School, Monterey, CA, Feb. 2010.

[45] J. S. Hesthaven, S. Gottlieb, and D. Gottlieb, *Spectral Methods for Time-Dependent Problems*. Cambridge: Cambridge University Press, 2007.

[46] C. Canuto, M. Y. Hussaini, A. Quarteroni, and T. A. Zang, *Spectral Spectral Methods, Fundamentals in Single Domains*. Berlin Heidelberg: Springer-Verlag, 2006.

[47] J. S. Hesthaven and T. Warburton, *Nodal Discontinuous Galerkin Methods, Algorithms, Analysis, and Application*. New York: Springer Science + Business Media, LLC, 2000.

[48] W. Gautschi, *Lecture notes in pure and applied mathematics: asymptotics and computational analysis*, vol. 124, ch. How (Un)stable are Vandermonde systems?, pp. 193–210. New York, NY: Marcel Dekker, Inc., 1990.

[49] P. Erdos, “Problems and results on the theory of interpolation,” *Acta Math. Acad. Sci. Hungar.*, vol. 12, pp. 235–244, 1961.

[50] S. Leon, *Linear Algebra*. 8th ed. Upper Saddle River, NJ: Pearson Prentice Hall, 2010.

[51] J. S. Hesthaven, "From electrostatics to almost optimal nodal sets for polynomial interpolation in a simplex," *SIAM Journal on Numerical Analysis*, vol. 35, pp. 655–676, Apr. 1998.

[52] B. Fornberg, *A Practical Guide to Pseudospectral Methods*. No. 1 in Monographs on Applied and Computational Series, Cambridge: Cambridge University Press, 1995.

[53] L. N. Trefethen and J. A. C. Weideman, "Two results on polynomial interpolation in equally spaced points," *Journal of Approximation Theory*, vol. 65, pp. 247–260, Jun. 1991.

[54] R. L. Burden and J. D. Faires, *Numerical Analysis*. 9th ed. Boston, MA: Brooks/Cole, Cengage Learning, 2011.

[55] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*. No. 55 in National Bureau of Standards Applied Mathematics Series, 10th ed. Washington, DC: U.S. Government Printing Offices, 1972.

[56] L. N. Trefethen, *Spectral Methods in MATLAB*. Philadelphia: SIAM, 2000.

[57] J. Shen, T. Tang, and L.-L. Wang, *Spectral Methods: Algorithms, Analysis and Applications*. No. 41 in Springer Series in Computational Mathematics, Berlin Heidelberg: Springer-Verlag, 2011.

[58] J. Boyd, *Chebyshev and Fourier Spectral Methods*. Mineola, NY: Dover Publications, 2000.

[59] G. Karniadakis and S. Sherwin, *Spectral/hp Element Methods for Computational Fluid Dynamics*. 2nd ed. New York: Oxford University Press, 2005.

[60] D. Estep and D. French, "Global error control for the continuous galerkin finite element method for ordinary differential equations," *RAIRO*, vol. 28, pp. 815–852, Nov. 1994.

[61] P. LeSaint and P. A. Raviart, *Mathematical aspects of finite elements in partial differential equations*, ch. On the finite element method for solving the neutron transport equation, pp. 89–145. New York: Academic Press, 1974.

[62] K. Bottcher and R. Rannacher, "Adaptive error control in solving ordinary differential equations by the discontinuous galerkin method," technical report, Institute of Applied Mathematics, University of Heidelberg, Heidelberg, Germany, Feb. 1997.

[63] A. Logg, "Multi-adaptive galerkin methods for odes, 1," *SIAM Journal on Scientific Computing*, vol. 24, pp. 1879–1902, May 2003.

[64] A. Logg, "Multi-adaptive galerkin methods for odes, 2: Implementation and applications," *SIAM Journal on Scientific Computing*, vol. 25, pp. 1119–1141, Dec. 2003.

[65] A. Logg, “Multi-adaptive galerkin methods for odes, 3: A priori error estimates,” *SIAM Journal on Numerical Analysis*, vol. 43, pp. 2624–2646, Jan. 2006.

[66] M. Delfour, W. Hager, and F. Trochu, “Discontinuous galerkin methods for ordinary differential equations,” *Mathematics of Computation*, vol. 36, pp. 455–473, Apr. 1981.

[67] W. Kang, Q. Gong, and I. M. Ross, “On the convergence of nonlinear optimal control using pseudospectral methhods for feedback linearizable systems,” *International Journal of Robust and Nonlinear Control*, vol. 17, pp. 1251–1277, Sep. 2007.

[68] W. Kang, I. M. Ross, and Q. Gong, *Analysis and design of nonlinear control systems - In honor of Alberto Isidori*, ch. Pseudospectral optimal control and its convergence theorems, pp. 109–124. Berlin Heidelberg: Springer-Verlag, 2008.

[69] P. Williams, “Optimal control of electrodynamic tether orbit transfer using timescale separation,” *Journal of Guidance, Control, and Dynamics*, vol. 33, pp. 88–98, Jan.-Feb. 2010.

[70] P. N. Desai and B. A. Conway, “Two-timescale discretization scheme for collocation,” *Journal of Guidance, Control, and Dynamics*, vol. 31, pp. 1316–1322, Sep.-Oct. 2008.

[71] Q. Gong, I. M. Ross, and W. Kang, “Some early results on multiscale pseudospectral optimal control,” in *Technical Report, 2012 AFOSR Dynamics and Control Program Review Meeting*, (Washington D.C.), AFOSR Dynamics and Control Program, Aug. 2012.

[72] G. Sansone and E. H. Diamond, *Orthogonal Functions*, vol. 9 of *Pure and Applied Mathematics*. New York: Interscience Publisher, Inc., 1959.

[73] R. Boucher, W. Kang, and Q. Gong, “Galerkin optimal control for constrained nonlinear problems,” in *Proceedings of the 2014 American Control Conference*, no. ThA18.4, (Portland, Oregon), IEEE, Jun. 2014.

[74] R. Boucher, W. Kang, and Q. Gong, “Discontinuous galerkin optimal control for constrained nonlinear problems,” in *Proceedings of the 2014 International Conference on Control and Automation*, no. WeC3.5, (Taichung, Taiwan), IEEE, Jun. 2014.

[75] K. A. Ross, *Elementary Analysis: The Theory of Calculus*. New York: Springer-Verlag, 2003.

[76] G. Bachman and L. Narici, *Functional Analysis*. New York: Academic Press, 1966.

[77] D. Gottlieb and S. A. Orszag, *Numerical Analysis of Spectral Methods: Theory and Applications*, vol. 26 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Philadelphia: Society for Industrial and Applied Mathematics, 1977.

[78] D. Gottlieb and C.-W. Shu, “On the gibbs phenomenon and its resolution,” *SIAM Review*, vol. 39, pp. 644–668, Dec. 1997.

[79] S. M. Kaber, “The gibbs phenomenon for jacobi expansions,” Tech. Rep. R05003, Laboratoire Jacques-Louis Lions, Universite Paris, Paris, France, Oct. 2005.

[80] J. R. Higgins, *Completeness and Basis Properties of Sets of Special Functions*. Cambridge: Cambridge University Press, 1977.

[81] D. Funaro, *Polynomial Approximation of Differential Equations*. No. 8 in Lecture Notes in Physics Monographs, New York: Springer-Verlag, 1992.

[82] G. Freud, *Orthogonal Functions*. Mineola, NY: Pergamon Press, 1971.

[83] R. Boucher, W. Kang, and Q. Gong, “Feasibility of the galerkin optimal control method,” in *Proceedings of the 2014 IEEE Conference on Decision and Control*, no. WeC12.5, (Los Angeles, California), IEEE, Dec. 2014.

[84] I. M. Ross, W. Kang, and Q. Gong, “Some recent results on pseudospectral optimal control,” in *Technical Report, 2009 AFOSR Dynamics and Control Program Review Meeting*, (Washington D.C.), AFOSR Dynamics and Control Program, Aug. 2009.

[85] Q. Gong, I. M. Ross, and F. Fahroo, “Costate estimation by a chebyshev pseudospectral method,” *Journal of Guidance, Control, and Dynamics*, vol. 33, pp. 623–628, Mar.-Apr. 2010.

[86] Q. Gong, F. Fahroo, and I. M. Ross, “Spectral algorithm for pseudospectral methods in optimal control,” *Journal of Guidance, Control, and Dynamics*, vol. 31, pp. 460–471, May-Jun. 2008.

[87] R. A. Adams and J. F. Fournier, *Sobolev Spaces*, vol. 140 of *Pure and Applied Mathematics Series*. Oxford: Elsevier Science Ltd, 2nd ed., 2003.

[88] E. D. Nezza, G. Palatucci, and E. Valdinoci, “Hitchhiker’s guide to the fractional sobolev spaces,” *Bulletin des Sciences Mathematiques*, vol. 136, pp. 521–573, Jul.-Aug. 2012.

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California